

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



BRAIN DRAIN

Cognitive decline linked to impaired waste removal from cerebral fluids by meningeal lymphatics **PAGES 172 & 185**

ENVIRONMENT

TROUBLED WATERS

How to turn the tide of oil-tanker accidents

PAGE 161

MATERIALS SCIENCE

WIRED UP FOR WEARING

Made-to-measure diode fibres for use in fabrics

PAGES 170 & 214

IMMUNOLOGY

INFLAMMASOME ACTIVATION

Mitochondrial DNA synthesis helps drive innate immunity

PAGES 176 & 198

NATURE.COM/NATURE

9 August 2018 £10

Vol. 560, No. 7717



THIS WEEK

EDITORIALS

WORLD VIEW A radical experiment could change how scientists are funded **p.143**

PENGUIN PLIGHT World's second-largest penguin colony has collapsed **p.144**



ISLAND GENES 'Hobbit' stature evolved twice in humans of Flores **p.145**

Signs of hope against influenza

A universal vaccine might be some way off, but research into how the immune system responds to the virus will be crucial to achieving that goal.

Hundreds of thousands of people die every year from influenza or a related condition. Unlike chicken pox and many other viruses in which the initial infection often provides lifelong immunity, the flu virus is constantly evolving. As a result, although most children will have been infected by the time they are around 3 years old, they will encounter new bouts of flu every 5–10 years.

As we highlight in a News Feature this week (see page 158), there is increasing scientific interest in how the human immune system is primed by its first exposure to flu in childhood. This immunologic 'imprinting' in part explains the vastly divergent susceptibilities of people born in different years to seasonal outbreaks of flu. The closer the characteristics of the circulating virus to the strain a person first experienced, the stronger their natural defences against it.

Such insights could help researchers to design more-effective vaccines. The need is great. In a good year, seasonal vaccines might protect six in ten people from infection, and this protection begins to wane after a few months. By contrast, a single shot of a vaccine against yellow fever is more than 99% effective and confers lifelong protection.

In the 2017–18 flu season, a virus subtype called H3N2 has predominated in many countries. In those places, the efficacy of the available vaccine has been much worse: just 10% in Australia, 17% in Canada and 25% in the United States. That's better than nothing, and is surely saving many precious lives — especially young children, who are particularly vulnerable. But it's hardly optimal.

The US Centers for Disease Control and Prevention notes that 80% of the children who died from flu this year in the United States had not been vaccinated. At present, many adults don't get immunized. A more-effective, longer-lasting vaccine would increase uptake, and would also probably reach the threshold to achieve collective — herd — immunity, thus reducing the number of people who could pass on the infection.

This logic is all the more important when it comes to the many lower- and middle-income countries that have few, if any, flu-vaccination programmes, simply because the costs and logistics of re-formulating and re-administering vaccines every year are prohibitive. Less-onerous vaccine requirements would encourage these countries to vaccinate their populations — and help to reduce the 290,000–650,000 deaths estimated by the World Health Organization to occur worldwide every year from flu-related respiratory diseases. Better still would be universal flu vaccines that would confer immunity against all new flu subtypes that emerge to cause pandemics.

Many scientists are increasingly confident that such a breakthrough is in reach. Are they right? Research over the past decade suggests that developing such vaccines is doable. This work includes findings that childhood imprints 'memorize' regions of the virus that mutate little and so differ little between flu subtypes. This memory is dormant, but it is there; a vaccine that could 'wake up' this memory should therefore produce broadly reactive antibodies that protect against multiple flu strains and subtypes. Technological advances, such as single-cell sorting and sequencing, are revolutionizing scientists' ability to characterize in great

depth the function of cell types involved in the host immune response.

Much essential information is expected to flow from a new large cohort study funded by the US National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland. Starting next year, it will monitor newborns from several countries over multiple flu seasons to see how their first flu infections, and any subsequent ones (and vaccinations), affect their immune systems. It will also chart how they respond to new exposures, and so help to unravel the mysteries of the mechanisms of their immunologic imprinting.

Commendably, the NIAID has stipulated that all data and clinical samples be shared widely with other researchers. This is important, because not only are such large and intensive cohort studies expensive and hard to run, but this study is of infants — meaning it's crucial to minimize blood draws and make the most of each sample.

We still have a way to go. But protective and longer-lasting vaccines seem to be well on their way. ■

"Universal flu vaccines would confer immunity against all new flu subtypes."

How I wonder

A star that hides its shine draws admiring looks.

Given that they spend their nights drenched in the astounding wonders of the cosmos, the word 'exceptional' probably comes a little harder for an astronomer than for those of us more concerned with the routine of Earthly pursuits. What can make a star-watcher draw breath and look again? A star with its coat on back to front would probably do it.

So it proves. In this week's *Nature Astronomy*, researchers detail their amazement at the newly discovered secrets of HuBi 1: a star that hides its shine beneath a murky shell of dust. Or, as the scientists put it: the first inside-out planetary nebula around a born-again star (M. A. Guerrero *et al.* *Nature Astron.* <https://doi.org/10.1038/s41550-018-0551-8>; 2018).

Typically, stars inside a planetary nebula ionize gaseous material previously ejected — so the surrounding shell of material closest to the star's surface is affected the most. But not HuBi 1. Here, the innermost regions are less ionized.

Simulations of stellar evolution suggest a likely — but rare — cause: the star had started to ionize its nebula, but then went through a period of rebirth to briefly flare again, re-igniting its nuclear fuel. In the process, it burped out a little extra material. This generated a shock wave that did some ionization of its own, but farther away from the surface. That shock wave is leaving dust behind as the material cools. Exceptional. ■



Who would you share your funding with?

I want to see whether the wisdom of crowds does a better job than conventional grant review at supporting research, says Johan Bollen.

Lawmakers in Chicago, Illinois, are readying a bill to test universal basic income. The idea is to give a fixed stipend to families without requiring reams of paperwork to assess eligibility. Advocates think that this streamlined system will allocate resources more fairly and with less bureaucratic bloat.

I propose that something similar could be used to fund science. In such a system, all qualified scientists would get some guaranteed funding — no grants required. But there should be one added step: everyone must anonymously allocate a fraction of their funds to other researchers of their own choosing.

The goal of this system would be to let scientists devote more of their time to research. The European University Association in 2016 estimated that the equivalent of at least one-quarter of Europe's Horizon 2020 funding programme goes to preparing grant applications (see go.nature.com/2vx3mjx). A 2013 study estimated that Australian scientists collectively spent more than five centuries of time preparing 3,727 proposals in 2012 (D. L. Herbert, A. G. Barnett and N. Graves *Nature* **495**, 314; 2013). Reviews might improve the quality of projects that are actually funded, but at what cost?

The scientific community is exploring ways of improving grant review, such as new evaluation systems or, as in New Zealand's Health Research Council, a modified lottery for promising proposals deemed both transformative and viable. But none of these substantially shrinks the bureaucratic burden. With current funding rates, researchers will continue to spend more time applying for grants with less-certain outcomes. That means less time doing science.

It is time to try something radical. I have spent the past five years trying to work out a crowd-based system, together with several colleagues. We call it Self-Organizing Funding Allocation (SOFA). Earlier this year, the Netherlands Organisation for Scientific Research held a workshop to plan a pilot test with SOFA, after the Dutch Parliament directed it to explore alternative modes of funding. Experts at the workshop agreed that the pilot project must be large enough — and last long enough — to make evaluation possible. We hope to publish evidence for, and a pathway to implement, this system within the next two years.

In SOFA, every participant starts with the same allocation of funding every year but must allot a portion to other scientists. Reasons to select someone could range from, 'That was a great paper' to 'I think they will release useful data.' Those who get the most give the most, because scientists give a percentage of everything received under SOFA. To avoid currying favour, this process will be anonymous.

Those who receive no donations still have their baseline. The 'baseline' and 'donation' cycles repeat every year. The distribution of

funding will reflect community consensus as to who deserves it.

SOFA retains the assumption at the heart of grant review that scientists know best who does good science, but it extends the process to all scientists instead of small review panels, and ensures a stable source of funding for early-career researchers. Funders can still develop grant programmes to encourage certain areas of research, such as neglected diseases or promising, risky new topics.

My team at Indiana University Bloomington ran a simulation assuming that all scientists funded by the US National Institutes of Health and the US National Science Foundation would donate to those they cited (J. Bollen *et al. EMBO Rep.* **15**, 131–133; 2014). This analysis of more than 100,000 investigators, 37 million papers and 770 million references yielded a hypothetical funding distribution surprisingly similar to that produced by grant review — without anyone submitting or reading a single application.

Of course, people don't always behave as predicted, as shown by the Brexit vote and the US 2016 presidential election. And freeing funding from proposals risks unleashing more sexism, racism and ableism than we already have.

We plan to build in precautions. We can limit collusions and kickback schemes — the financial equivalent of citation cartels — by mandating a minimum number of recipients and restricting people from designating frequent collaborators, or colleagues at the same institution. Counteracting gender, age and prestige biases that plague conventional peer review might even be easier in SOFA because they are measurable. Parameters

can be tuned to distribute funds according to desired criteria, for example, limiting repeated allocations to single institutions or individuals, and guaranteeing donations to under-represented groups.

Funders will need to define who gets to participate; perhaps everyone on a research track who is at an accredited institution and receiving a minimum salary. Otherwise, universities might be tempted to mint more professors and research associates. Also, without review panels, universities will need to be proactive to ensure that experiments fall within ethical guidelines and that scientists follow rules and fulfil obligations.

I understand scepticism that SOFA might not fund the highest-quality research: that friendship or flash might get in the way. But writing grant applications has already got in the way of doing research, and we owe it to science to find out whether this will work. The conventional proposal-based grant system might never have got off the ground had its adoption required the same level of proof we now seek. ■

Johan Bollen is a professor at the Indiana University Bloomington School of Informatics and Computing, Bloomington, Indiana, USA. e-mail: jbollen@indiana.edu

REVIEWS MIGHT
IMPROVE
PROJECTS
THAT ARE ACTUALLY
FUNDED,
BUT AT WHAT
COST?

SEVEN DAYS

The news in brief

POLICY

Genetic privacy

A group of DNA-testing companies has developed guidelines for sharing users' information with law-enforcement and for-profit companies. On 31 July, 23andMe in Mountain View, California, Ancestry in Lehi, Utah, and several other direct-to-consumer companies released a document describing standards for how genetic data are stored and used — including user privacy, data security and the legal process for sharing data with police. The guidelines also ban the sharing of genetic data with entities such as employers and insurance companies without the user's consent. The move comes after California investigators identified a suspect in a series of rapes and murders known as the Golden State Killer case by comparing DNA from crime scenes with genetic data that the suspect's relatives had submitted to the testing company GEDmatch. The privacy of genetic data also drew attention in July, when 23andMe announced that it would share user data, with permission, with GlaxoSmithKline after the pharmaceutical giant invested US\$300 million.

EVENTS

Ebola returns

Barely a week after the end of an outbreak of Ebola virus was declared in the northwest of the Democratic Republic of the Congo, a new outbreak has emerged in North Kivu province, some 2,500 kilometres to the east. The latest episode is potentially more dangerous, because it is in a war zone, which will complicate response efforts. Many thousands of people are fleeing from violent conflict to other regions

and to nearby nations such as Uganda, Rwanda and Burundi. So far, 16 cases of Ebola have been confirmed in the latest outbreak, and the virus is thought to have caused more than 30 deaths. Confirmation of the virus strain's identity is expected this week; it is suspected to be the EBOV strain, for which an experimental vaccine is available. However, the movement of people will complicate the 'ring vaccination' strategy used in previous outbreaks, because this depends on rapidly immunizing the contacts of infected people and their contacts, as well as front-line responders. "We are at the top of the degree of difficulty scale

in terms of responding to this outbreak," said Peter Salama, head of the World Health Organization's emergencies programme, on 3 August.

FUNDING

Grants cancelled

The March of Dimes foundation, a US non-profit group focused on improving child health, has abruptly terminated 37 research grants totalling US\$3 million. On 24 July, grant recipients received an e-mail from the foundation in White Plains, New York, informing them that their three-year grants had been cut off retroactively, starting on 30 June. The foundation made the decision

to revoke the grants because of a budget shortfall. Going forward, the group's board of directors has decided to restrict its research support to studies on reducing premature births, said Kelle Moley, the March of Dimes' chief scientific officer. The organization will continue to fund young investigators through its prestigious Basil O'Connor awards.

UNIVERSITIES

Power imbalance

Academics should be removed from their supervisory, teaching or assessment roles if they develop a romantic relationship with one of their students, say organizations that represent Australian



CHIME COLLABORATION

Canadian telescope spots cosmic burst

A radio telescope, inaugurated last year, has detected its first fast radio burst (FRB), giving astronomers a powerful weapon for studying these mysterious events. The 2-millisecond signal, announced on 1 August, heralds an expected deluge for the Canadian Hydrogen Intensity Mapping Experiment (CHIME): once fully operational, CHIME should record more than a dozen FRBs a day. Astrophysicists have

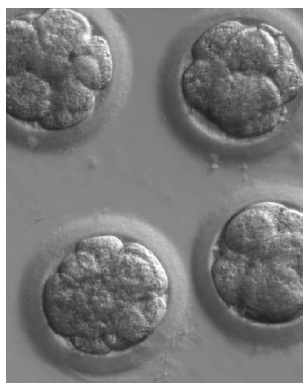
proposed a number of explanations for these events — which are fast bursts of radio energy — from evaporating black holes to erupting neutron stars, but data have been scarce so far. CHIME consists of 4 reflectors shaped like half-pipes, each 100 metres long. Its primary science goal is to map the density of interstellar hydrogen across the Universe in the epoch between 10 billion and 8 billion years ago.

universities and students. Guidelines released on 1 August state that romantic relationships between academic supervisors and their students are never appropriate because they create a power imbalance. Although some institutions have their own policies in place for addressing such situations, Australia is one of the first countries to issue national principles. They were developed by four organizations that represent the country's universities, academics and postgraduate students. The guidelines were created in response to a 2017 survey of more than 30,000 Australian university students about their experiences of sexual harassment or sexual assault. It found that postgraduate students were almost twice as likely as undergraduates to have been sexually harassed by a lecturer or tutor. See go.nature.com/2ocm48b for more.

RESEARCH

Embryo editing

The authors of a controversial study — which used gene editing in human embryos in an effort to fix a disease mutation — have responded to criticisms of their work in an 8 August communication



to *Nature*. In the original study, published last August (H. Ma *et al.* *Nature* **548**, 413–419; 2017), a team led by reproductive biologist Shoukhrat Mitalipov at Oregon Health & Science University in Portland described using CRISPR–Cas9 gene editing to correct a mutation, which causes a heart condition, in human embryos. The embryos (**pictured**) were created from sperm that carried the mutation and healthy donor eggs; Mitalipov's team reported that the sperm's faulty version of the gene was replaced with a copy from the egg during the gene-editing process. Critics of the work — including authors who have formally published their responses in two separate reports in *Nature* this week — say that Mitalipov's

team had not ruled out alternative explanations for its results. In its response, Mitalipov's team reports new data it says back up its claims that gene correction occurred in the manner proposed.

PEOPLE

Scientists speak out

The US Society for Neuroscience and the Federation of European Neuroscience Societies published a joint statement on 3 August criticizing the Max Planck Society (MPS) in Germany for its treatment of neuroscientist Nikos Logothetis, who used to run a primate laboratory. A director at the Max Planck Institute for Biological Cybernetics in Tübingen, Logothetis is charged with mistreatment of animals following allegations by animal-rights groups. A court has not ruled on the charges, which Logothetis denies. But the MPS has removed many of his responsibilities relating to animal research. The neuroscience societies, which together represent more than 60,000 scientists, say the MPS's treatment of Logothetis sets an “alarming precedent” that disregards the presumption of innocence. MPS president Martin Stratman declined to

comment on the statement, but in response to earlier criticism of MPS's handling of the affair, he said that the society restricted Logothetis's work to reassure the public that it takes animal welfare seriously.

Krauss departs

Physicist Lawrence Krauss, a prominent sceptic and writer, is leaving his post as director of the Origins Project at Arizona State University (ASU) in Tempe, a multidisciplinary centre that he founded nine years ago. In February, BuzzFeed News reported a number of allegations of sexual harassment against Krauss. The university placed him on paid leave in March while it investigated the reports. Krauss, who has previously denied the allegations, declined *Nature's* request for comment. In a 2 August tweet, Krauss said that the university had decided not to renew his appointment as director when it expired in July. ASU confirmed the decision but declined to comment further. Krauss's successor will be Lindy Elkins-Tanton, a planetary scientist at ASU. The Origins Project holds workshops, lectures and other discussions aimed at exploring the origins of the Universe, life, consciousness and more.

TREND WATCH

An Australian city has seen local cases of dengue fever plunge after it was blanketed with mosquitoes modified to block transmission of the virus. Over 28 months beginning in October 2014, researchers and community members released roughly 4 million *Aedes aegypti* mosquitoes over 66 square kilometres in Townsville. The insects carried *Wolbachia* bacteria that block them from transmitting dengue, Zika and some other disease-causing viruses.

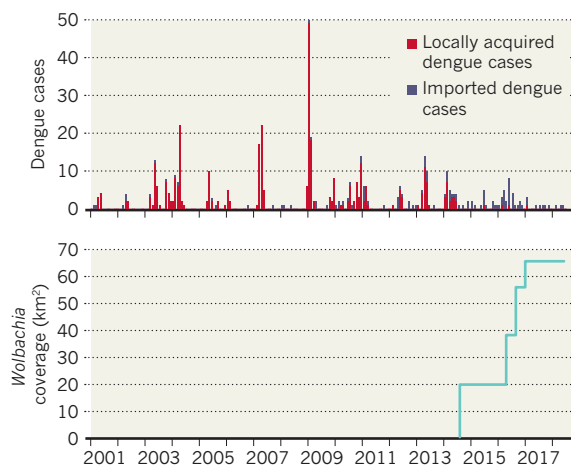
A team led by microbiologist Scott O'Neill at Monash University in Clayton, Australia, tracked the mosquito release

— the first time the strategy has been trialled across an entire city. *Wolbachia*-infected mosquitoes quickly spread the bacteria to local mosquito populations. In many suburbs, nearly 100% of mosquitoes carried *Wolbachia* one year after the release period.

Townsville, which has a population of around 187,000, has faced periodic dengue outbreaks since 2001. In the 44 months after the releases began, however, authorities recorded just 4 locally acquired dengue cases, compared with 54 locally acquired cases over the 44 preceding months. (During the same period after the release, 51 imported cases were reported.)

MOSQUITO TOWN

After the release of around 4 million mosquitoes — all carrying bacteria that stop the insects transmitting dengue virus — locally acquired cases of dengue plummeted in Townsville, Australia.



NEWS IN FOCUS

POLITICS Trump nominates a meteorologist to be his science adviser **p.150**

PHYSICS Thousands of exotic topological materials emerge from systematic hunt **p.151**

POLICY Gigantic review of German science finds lack of diversity **p.153**



INFLUENZA Why the immune system's first impressions mean so much **p.158**

RITESH SHUKLA/NURPHOTO/GETTY



The Ganges is one of the world's most polluted rivers.

GEOGRAPHY

Indian scientists race to map Ganges river in 3D

Digital models of the river and settlements will help authorities to reduce pollution.

BY LOU DEL BELLO

Scientists and engineers are about to begin the monumental task of mapping the vast stretch of the Ganges river that runs through India, in unprecedented detail. They hope to get started on the work before the monsoon brings bad weather that could delay the project.

Their goal is to create the most comprehensive picture yet of the topography of the river and the human settlements that surround it, to

track sources of waste and help authorities clean up one of the world's most polluted waterways.

"It's a race against time," says Girish Kumar, who heads the national surveying agency, the Survey of India based in Dehradun in the Himalayan foothills, which is leading the project.

Although the mapping is expected to take about eight months, the team is eager to get started in case the monsoon season, which began in June, forces them to ground the planes that will be doing much of the work.

A fleet of small aircraft equipped with lidar

instruments will soon start scanning the 2,525-kilometre stretch of river that passes through five Indian states — one metre at a time. Lidar is a technique similar to radar, in which instruments bounce laser pulses off the ground. The researchers will use it to produce digital elevation models of the watercourse and the hundreds of thousands of buildings that sit up to 10 kilometres either side of the riverbank.

If the schedule goes to plan, the 3D maps should be available by the end of next year.

The project will produce high-resolution ►

► maps of the drainage systems of major cities along the Ganges — the network of discharge outlets that release sewage and commercial waste water into the river. An estimated 600 million people live in the Ganges basin, and rely on water from the river for drinking and bathing. The Ganges is sacred to the country's large Hindu population, who view the river as an embodiment of the goddess Ganga and use its waters in religious rituals.

Although some sources of waste in the Ganges are well known, detailed models of how pollution enters and moves along the river will enable officials to design more-effective reduction strategies. Environmental engineer Vinod Tare of the Indian Institute of Technology in Kanpur says that many current government interventions, such as diverting raw industrial sewage away from the river, are implemented without sufficient information to assess whether they are working. "Right now, we do not even have a simple topography of the basin," says Tare, who has been involved in Ganges-management research for more than three decades.

Government officials also hope to use the maps to improve understanding of how cities develop along the riverbank, and of how the bank is being eroded. This will help local governments to manage risks such as floods. "We will have a better idea of what industries



and human settlements will be most affected," says Kumar.

The mapping project (see 'Mapping Mother Ganges') will cost 870 million rupees (US\$12.7 million). "It is expensive, but compared to what we will be spending to address the pollution problem, it is hardly anything," says Tare.

But water-quality researcher Abed Hossain says the benefits of detailed monitoring will go unrealized if researchers cannot access all the information and use it to develop models and interventions. If the mapping doesn't go

as planned, the government could become worried about negative publicity and restrict access to some of the raw data, says Hossain, who works at the Bangladesh University of Engineering and Technology in Dhaka. In south Asia, he says, "governments are edgy about failures".

Kumar says that the government has issued guidelines for data sharing and will share the information collected for the project.

The mapping is part of the Indian government's renewed push to use technology to monitor and clean the Ganges. In 2015, the government approved the 200-billion-rupee National Mission for Clean Ganga, a wide-ranging effort that includes improving the treatment of sewage and reducing industrial pollution.

But as the deadline of 2020 approaches, the government is still a long way from meeting many of its targets. Last year, the independent auditor-general found that the clean-up effort had been delayed by financial mismanagement and poor planning and implementation.

The management of the river is shaping up to be a central issue in the lead-up to the general election next year. Kumar says that the maps will be a crucial resource for future interventions. "Before planning anything, we need a map," he says. ■

POLICY

Trump finally nominates a science adviser

Meteorologist Kelvin Droegemeier would lead the White House science office.

BY SARA REARDON & ALEXANDRA WITZE

US President Donald Trump has nominated meteorologist Kelvin Droegemeier as his government's top scientist. If confirmed by the Senate, Droegemeier would lead the White House Office of Science and Technology Policy (OSTP).

Trump, who took office 19 months ago, has gone longer without a top science adviser than has any first-term president since at least 1976. He announced his pick on 31 July.

"My initial reaction is, wow, they found someone," says Kei Koizumi, visiting scholar at the American Association for the Advancement of Science in Washington DC and a former assistant director at the OSTP under president Barack Obama.

Droegemeier would be the first non-physicist to serve as White House science

adviser since Congress established the OSTP in 1976. "I think he is a very solid choice," says John Holdren, who led the OSTP for eight years as Obama's science adviser. "He is a respected senior scientist and he has experience in speaking science to power."

An expert on extreme-weather events, Droegemeier has been vice-president for research at the University of Oklahoma in Norman since 2009. Last year, Oklahoma Governor Mary Fallin, a Republican, appointed him as the state's secretary of science and technology. The meteorologist has also served on the National Science Board (NSB), which oversees the National Science Foundation, under presidents Obama and George W. Bush. Droegemeier led NSB committees on hurricane science and research administration, among other topics, and was the board's vice-chairman from 2012 to 2016.

"He combines a lot of qualities in somebody

you'd like to see in public service," says Roger Pielke Jr, a political scientist at the University of Colorado Boulder who has studied the history of US science advisers and who worked with Droegemeier in the 1990s and early 2000s. "He is, in the most positive way, a nerdy meteorologist who loved working on weather technology. And he also has a knack for administration and working his way around the system."

If confirmed, Droegemeier will take control of an office radically reshaped by the Trump administration. The president has reduced the number of OSTP staff members to about 50, well below the 130 employed by Obama. The Trump team has also placed greater emphasis on technology issues, and has repeatedly sought to cut or eliminate high-profile science programmes — including a public-health-preparedness fund at the Centers for Disease Control and Prevention, climate-change

AP/SHUTTERSTOCK

programmes at the Environmental Protection Agency and NASA's Wide-Field Infrared Survey Telescope.

PLAYING CATCH-UP

Some of Droegemeier's colleagues hope that he would help to shift the Trump administration's thinking on climate change. "I'm certain he believes in mainstream climate science," says Rosina Bierbaum, an environmental-policy expert at the University of Michigan in Ann Arbor who has held multiple presidential-advisory roles. Bierbaum and Droegemeier worked on climate-change issues together while on the board of the University Corporation for Atmospheric Research in Boulder, Colorado. "He's an excellent communicator and really good at distilling complex issues," she says.

The OSTP has managed to keep working without a permanent director, developing strategies to monitor space weather and boost science, technology, engineering and mathematics education. But Koizumi says that Trump would benefit from having a science adviser to consult when making decisions on issues such as natural disasters.

Because the position has remained vacant for so long, "they'll be filling from behind" to get the OSTP fully staffed, says Phil Larson, a senior adviser in Obama's OSTP who is now assistant dean of engineering at the University of Colorado Boulder. "Now the question will be, will his voice be represented around the table in the discussions that are going on at the highest levels of the government?"

And serving as the top scientist in an administration that has been criticized for its science policy could be difficult in other ways. "Droegemeier's going to get all sorts of



Kelvin Droegemeier is an expert in extreme-weather events.

questions," says Pielke. "There's going to be a tremendous amount of pressure." He sees a probable analogue in the experiences of John Marburger, the physicist who advised president George W. Bush.

Marburger was sharply criticized for supporting government policies that were unpopular with the scientific community — such as Bush's decision to withdraw from the Kyoto Protocol on climate change and to restrict federally funded research on embryonic stem cells. "It's going to get tough pretty

quickly for [Droegemeier]," Pielke says.

It is not clear whether the White House intends to appoint Droegemeier as an assistant to the president, a position held by several recent White House science advisers — including Holdren. The title, which is separate from that of OSTP director, essentially signals close ties to the president and his top aides. An OSTP spokesperson says that any decision about whether to give Droegemeier an additional title would be made after his confirmation by the Senate. ■

TOPOLOGY

Trove of exotic matter thrills physicists

Thousands of new 'topological' materials are emerging as researchers exploit new algorithms to scour databases.

BY ELIZABETH GIBNEY

The already buzzing field of topological physics could be about to explode. For the first time, researchers have systematically scoured entire databases of materials in search of ones that harbour topological states — exotic phases of matter that have fascinated physicists for a decade. The results show that thousands of known materials probably have topological properties — and perhaps up to 24%

of materials in all. Previously, researchers knew of just a few hundred topological materials, and only around a dozen have been studied in detail.

"I'm shocked by the number," says Reyes Calvo, an experimental physicist at the nanogUNE Cooperative Research Center in San Sebastián, Spain.

In July, several teams posted preprints^{1,2,3} online detailing their scans of tens of thousands of materials and their predicted topological classifications, which are based on algorithms that

use a material's chemistry and symmetry to calculate their likely properties. Two teams have already integrated their algorithms into searchable databases. "You can put in a compound name and, with one click, get whether there is topology or not. For me, this is wonderful," says Chandra Shekhar, a condensed-matter physicist at the Max Planck Institute for Chemical Physics of Solids in Dresden, Germany.

The resulting haul of topological materials could bring scientists closer to practical applications for these exotic phases, which could revolutionize electronics and catalysis. "The more materials with unusual properties we know, the more chance there will be of a breakthrough," says Oleg Yazyev, a physicist at the Swiss Federal Institute of Technology in Lausanne.

These materials derive their unusual features from their topology. In mathematics, topology is the study of objects with properties that remain unchanged when they are smoothly deformed and not torn. In materials, topology applies not to the shape of a solid object, but to the geometry of an abstract description of its electrons' quantum states. Their topology ►

► nature means these states are resistant to change, and thus stable to temperature fluctuations and physical distortion — features that could make them useful in devices.

Physicists have been investigating one class, known as topological insulators, since it was first seen experimentally⁴ in 2008. Topological insulators consist mostly of insulating material, yet their surfaces are great conductors. And because currents on the surface can be controlled using magnetic fields, physicists think the materials could find uses in energy-efficient ‘spintronic’ devices, which encode information in a kind of intrinsic magnetism of particles known as spin. But despite a decade of study, physicists have yet to find a topological insulator that has properties suitable for use in devices — for example, a material that is easy to grow, non-toxic and with tunable electronic states at room temperature.

The newly released catalogues classify all non-magnetic materials with known crystal structures by their topology, using methods published last year. Until now, physicists had largely relied on complex theoretical calculations to predict whether a specific material should harbour topological states. But in 2017,

Andrei Bernevig, a physicist at Princeton University in New Jersey, and Ashvin Vishwanath, at Harvard University in Cambridge, Massachusetts, separately pioneered approaches^{5,6} that speed up the process. The techniques use algorithms to sort materials automatically into

“It’s up to experimentalists to uncover new exciting physical phenomena.”

databases on the basis of their chemistry and properties that result from symmetries in their structure. The symmetries can be used to predict how electrons will behave, and so whether a material is likely to host topological states.

Applying Bernevig’s principles, a team led by researchers at the Beijing National Laboratory for Condensed Matter Physics scanned 39,519 materials and found more than 8,000 that are likely to have topological states. This includes both topological insulators and topological semimetals, which allow the study of new quantum phenomena and are being explored for use as catalysts. The team’s database is available for anyone to access and can be searched using a range of variables.

Bernevig and his colleagues also used their method to create a new topological catalogue. His team used the Inorganic Crystal Structure Database, filtering its 184,270 materials to find 5,797 “high-quality” topological materials. The researchers plan to add the ability to check a material’s topology, and certain related features, to the popular Bilbao Crystallographic Server. A third group — including Vishwanath — also found hundreds of topological materials.

Experimentalists have their work cut out. Researchers will be able to comb the databases to find new topological materials to explore. “We now have a large database of candidate materials, and it’s up to experimentalists to uncover new exciting physical phenomena,” Yazyev says. ■

1. Zhang, T. et al. Preprint at <https://arxiv.org/abs/1807.08756> (2018).
2. Vergniory, M. G., Elcoro, L., Felser, C., Bernevig, B. A. & Wang, Z. Preprint at <https://arxiv.org/abs/1807.10271> (2018).
3. Tang, F., Po, H. C., Vishwanath, A. & Wan, X. Preprint at <https://arxiv.org/abs/1807.09744> (2018).
4. Hsieh, D. et al. *Nature* **452**, 970–974 (2008).
5. Bradlyn, B. et al. *Nature* **547**, 298–305 (2017).
6. Po, H. C., Vishwanath, A. & Watanabe, H. *Nature Commun.* **8**, 50 (2017).

AWARDS

Number-theory prodigy among winners of coveted maths prize

Fields Medals awarded to researchers in number theory, geometry and differential equations.

BY DAVIDE CASTELVECCHI

Number theorist Peter Scholze, who became Germany’s youngest ever full professor at the age of 24, and geometrician Caucher Birkar — a Kurdish refugee — are among the winners of this year’s Fields Medals, the most coveted awards in mathematics. The medals, which are given out every four years, were presented on 1 August; the other recipients were Alessio Figalli, whose research involves differential equations, and Akshay Venkatesh, who also works on number theory. The winners’ names were announced in Rio de Janeiro, Brazil, at the opening of the International Congress of Mathematicians.

The Fields Medals, given out by the International Mathematical Union, are awarded to up to four mathematicians aged 40 or younger. For the first time in the medals’ 82-year history, none of the awardees are citizens of the United States or France — two countries that together have netted nearly half of the medals so far. Maryam Mirzakhani, a winner in 2014, remains the only woman ever to receive the prize. (Mirzakhani died of cancer in 2017.)



Fields medallists (left to right) Akshay Venkatesh, Peter Scholze, Alessio Figalli and Caucher Birkar.

Few observers doubted that Peter Scholze deserved a Fields Medal, or that he would win one this year. The 30-year-old became famous at 22 for finding a way to drastically shorten

a book-length proof in arithmetic geometry.

Scholze is now a professor at the University of Bonn in Germany, and a director at the Max Planck Institute for Mathematics in the

PABLO COSTA/ICM2018

same city. Most of his work has connections to ' p -adic fields', exotic extensions of the ordinary number system that are useful tools for studying prime numbers. On the p -adics, he has built fractal-like structures called perfectoid spaces, which have helped to solve problems across several fields of mathematics, including geometry and topology. In recent months, Scholze has been checking a gigantic proof of the abc conjecture, one of the biggest unsolved problems in number theory. In 2012, the enigmatic Japanese mathematician Shinichi Mochizuki posted a proof online, but no one has yet been able to say definitively whether it checks out. Now, Scholze and a colleague, Jacob Stix, are said to have found a significant gap in the proof.

Caucher Birkar, 40, has made breakthroughs in the classification of algebraic varieties — geometric objects that arise from polynomial equations, such as $y=x^2$. He was born in 1978 in a region of western Iran dominated by the Kurdish ethnic group. Birkar recalls his childhood in video profiles of the Fields medallists: "My parents are farmers, so I spent a huge amount of time actually doing farming," he says. "In many ways, it was not the ideal place for a kid to get interested in something like mathematics."

In 2000, after studying at the University of Tehran, Birkar moved to the United Kingdom, where he got refugee status and, eventually, UK citizenship. He is now a researcher at the University of Cambridge. Birkar said that he hopes that his Fields Medal will put "just a little smile on the lips" of the world's estimated 40 million Kurds.

His win made headlines for more than just his research: before the award ceremony was over, his briefcase was stolen, with his medal in it. The organizing committee of the congress presented him with a replacement medal in a special ceremony on 4 August.

Akshay Venkatesh, who is 36, works on, among other things, classical problems in number theory, including number systems that consist of fractions of whole numbers and roots such as $\sqrt{2}$. He is among the few mathematicians who have made substantial progress on a question formulated by mathematician Carl Friedrich Gauss in the nineteenth century. Venkatesh was born in New Delhi and raised in Australia, and is currently at the Institute for Advanced Study in Princeton, New Jersey.

Compared with the other three medallists, 34-year-old Alessio Figalli works in an area that is closer to the real world: optimal transport, which seeks the most efficient ways to distribute goods on a network. Figalli, who is Italian and works at the Swiss Federal Institute of Technology in Zurich, applies the field to partial differential equations, which have several variables and most often arise in physics. ■

POLICY

German science goes under a microscope

Gigantic review of Helmholtz centres finds lack of diversity.

BY QUIRIN SCHIERMEIER

Germany's largest research organization is funding top-notch science, but it needs to employ more foreign and female researchers — and it is failing to leverage 'big data', such as electronic medical records.

These are the conclusions emerging from a first-of-its-kind evaluation of the Helmholtz Association of German Research Centres, which employs some 30,000 scientists and technicians at 18 centres and has an annual budget of €4.5 billion (US\$5.3 billion).

Helmholtz showed *Nature* the results of the review, which individual centres will release over the next few weeks.

The results will serve as the basis for a strategic evaluation next year, which will be used to allocate research funding from 2021 to 2027. Other leading science organizations rarely, if ever, conduct such sweeping reviews, says neuroscientist Otmar Wiestler, president of the association.

DISCIPLINED ANALYSIS

"We were very impressed by the quality of the science," says Andrew Harrison, chief executive of the Diamond Light Source at the Harwell Science and Innovation Campus in Didcot, UK. He was one of more than 600 independent scientists from 27 countries who, between October 2017 and April 2018, spent up to a full week in Germany assessing the strengths and weaknesses of the organization's national research centres.

"As everywhere, the gender balance could be much better — but Helmholtz is aware of this and committed to improve it," adds Harrison.

In many fields — including biomedical research, condensed-matter physics and

materials sciences — Helmholtz centres rank among the world's top institutes by quality of basic science and research infrastructures, reviewers concluded. Energy research and Earth and environmental sciences also received high marks.

In biomedical research, reviewers endorsed the organization's current focus on infec-

"Like almost everywhere in science, real equity may still be generations away."

tious diseases, diabetes, dementia and cancer. But specialized health-research centres in Munich, Braunschweig, Bonn and Heidelberg must make better use

of patient data to develop diagnostic tools and therapies, the review concludes. It also recommends that the centres establish more designated clinical-trial units, in collaboration with hospitals, to take discoveries from the bench to practice.

"Reviewers have clearly seen that Germany is lagging behind in digital medicine," says Wiestler. "It is absolutely vital for health research and health care in this country that we catch up."

A CHALLENGE TO DO BETTER

Reviewers also urged the organization to boost diversity. Efforts to that effect are already under way, says Wiestler. A €5.4-million initiative to recruit more female scientists was launched last year. It aims to increase the proportion of women in senior positions, from the current level of 19% to 24% by 2020.

To attract more foreign scientists — the organization employs around 6,000 right now — Helmholtz plans to establish an international research school in astronomy, in partnership with the National University of ▶



**MORE
ONLINE**

STAY CURRENT

- Fury at US environmental agency's effort to weaken fuel standards go.nature.com/2jx6my4
- Scientists stunned as non-profit group ends research grants go.nature.com/2vjdhah
- Language-evolution theory questioned go.nature.com/2ohojjq

NATURE PODCAST

The immune system's effect on microbiota; geoengineering's impact on farming; and the genetics of fox aggression
nature.com/nature/podcast





This station in Antarctica is part of Helmholtz's Alfred Wegener Institute for Polar and Marine Research.

► General San Martín in Buenos Aires, and another in energy research, with five partner institutes in Israel. A prototype research school in cancer biology was recently launched in partnership with the Weizmann Institute of Science in Rehovot, Israel.

"I'm impressed by the seriousness of how Helmholtz is thinking about diversity and

gender equity," says reviewer Meigan Aronson, a condensed-matter physicist at Texas A&M University in College Station. "And yet, like almost everywhere in science, real equity may still be generations away."

Most Helmholtz centres also operate large research infrastructure, including light, ion and neutron sources; an experimental fusion

reactor; marine research vessels and aircraft; satellite systems; and Germany's Antarctic research station.

These facilities are Helmholtz's strongest asset, says Aronson, who spent a week last December helping to review neutron and nuclear research at the Helmholtz centre in Jülich.

Research time at these and other Helmholtz physics centres is in high demand. For example, an electron-positron collider called DESY, in Hamburg, and the synchrotron-radiation sources named BESSY in Berlin are used by scientists around the world to probe the structure of matter and experimental materials. Overall, almost 4,500 guest scientists spent time at Helmholtz centres in 2017. A €1.5-billion international accelerator facility for research with antiprotons and ions in Darmstadt, due to open around 2025, will add to Helmholtz's appeal, says Aronson.

The results of the extensive review will be analysed by the Helmholtz's leadership in the coming months. But already, says Harrison, the meticulously planned exercise has set a new standard for the evaluation of science.

"Reviewers are sometimes confronted with science organizations that don't completely engage with the process," he says. "Here, we went away with the feeling that every stone we could think of was turned over." ■



THE MICE THAT GROW HUMAN TUMOURS

They were supposed to be ideal models of disease. Now researchers are discovering the limits of patient-derived xenografts.

Lindsey Abel takes an anaesthetized mouse from a plastic container and lays it on the lab bench. With a syringe, she injects a slurry of pink cancer cells under the skin of the animal's right flank. These cells once belonged to a person with tongue cancer, a former smoker whose disease recurred despite radiation and surgery. The mouse is the second rodent to harbour them, creating a model for cancer known as a patient-derived xenograft (PDX). The tumour that grows inside will provide cells that can be transferred to more mice.

Abel has performed this procedure hundreds of times since she joined Randall Kimple's lab at the University of Wisconsin–Madison. Kimple,

BY CASSANDRA WILLYARD

a radiation oncologist, uses PDX mice to carry out experiments on human tumours that would be impractical in people, such as testing new drugs and identifying factors that predict a good response to treatment. His lab has created more than 50 PDX mice since 2011.

Kimple's lab is not the only one doing this; PDX mice have exploded in popularity over the past decade and are beginning to supplant other techniques for modelling cancer in research and drug development, such as mice implanted with cancer cell lines. Because the models use fresh human tumour fragments rather than

cells grown in a Petri dish, researchers have long hoped that PDXs would model tumour behaviour more accurately, and perhaps even help to guide treatment decisions for patients. They also allow researchers to explore the vast variety of human tumours. PDXFinder, a catalogue launched earlier this year, lists more than 1,900 types of PDX mice. But there are many more scurrying around in academic and industry labs — as many as 10,000 PDXs have been created, says Nathalie Conte, a bioinformatician at the European Bioinformatics Institute, in Hinxton, UK, who leads PDXFinder.

PDX models are not perfect, however — and scientists are beginning to recognize their

NIK SPENCER/NATURE

shortcomings and complexities. The tumours can diverge from the original sample, for example, and the models cannot be used to test immunotherapies. Now, biologists are scrutinizing PDX mice and looking for creative ways to cope with the challenges. “Every model is artificial in some way,” says Jeffrey Moscow, head of the investigational drug branch at the National Cancer Institute in Bethesda, Maryland. “The real question is how predictive are these models going to turn out to be.”

RISE AND FALL OF THE AVATARS

Scientists have been transplanting human cancers into mice for more than 50 years. In the 1960s, for example, researchers removed a tumour from a 74-year-old woman with colon cancer, minced it and injected the fragments under the skin of mice without immune systems. The tumours grew and were then cut up and transplanted into more mice. The approach didn’t gain much traction, however. Instead, many researchers relied on mice implanted with human cancer cells that had been grown in a dish, because that is cheaper and easier than using fresh tumour fragments from biopsies.

But in the early 2000s, researchers began to worry that cell-line xenograft models might not be very representative of human cancers. They realized that drugs that worked in these mice rarely worked as well in people, in part because the cells change in culture over time. So researchers turned again to PDX models.

One early adopter was Manuel Hidalgo, a cancer researcher at Harvard Medical School in Boston, Massachusetts. In 2002, he began working with a woman who had bile-duct cancer. Hidalgo proposed injecting her tumour cells directly into mice and seeing which drugs worked best on them. Four years later, Hidalgo co-founded a company aimed at generating these mouse ‘avatars’ for many more patients. That company — now part of Champions Oncology in Hackensack, New Jersey — began offering these models to oncologists and patients as a tool for determining the treatments most likely to work. Some people predicted that personalized mouse models would become a routine part of cancer treatment.

But the approach didn’t pan out the way the company had hoped, Hidalgo says. Last year, he and his colleagues published a study¹ that included 1,163 people who sought the services of Champions Oncology. Because not all tumours grow in mice, the company managed to generate PDX models for only half of them.

For many of these people, the mice came too late or physicians didn’t follow up with avatar testing. Still, the models do seem to be predictive: the researchers identified 92 patients who received treatments based on testing in the PDX models, and found that the PDX predictions were accurate 87% of the time.

Although the company still creates avatars for people who want them, it shifted its focus away from the personalized models about three years ago, according to chief executive Ronnie Morris.

They took too long to deliver answers, and they cost too much. “It was just a bad business for us,” Morris says.

SCIENTIFIC STAND-INS

Meanwhile, the popularity of PDX mice has soared in the research realm. Scientists have embraced the models to improve their understanding of tumour biology and to find new drugs. And yet questions remain as to whether they are better than previous models.

Todd Golub, head of the cancer programme at the Broad Institute in Cambridge, Massachusetts, and his colleagues analysed the genomes of hundreds of PDX models representing dozens of cancer types. They were looking at duplicated stretches in the genome and how they changed as the tumour cells passed through several live mice². The tumours evolved quickly: by the fourth passage, 88% of the PDX models had at least one large chromosomal aberration, and a median of 12% of the genome had been affected.

Juliet Williams, head of oncology pharmacology at Novartis in Cambridge, says it has been clear for some time that genetic changes occur. “The question is, does that small amount of drift that you see matter functionally?” she says. In 2015, Williams and her colleagues put together a panel of 250 PDX models and used them to

“EVERY MODEL IS ARTIFICIAL IN SOME WAY.”

test more than 60 drugs and drug combinations, including a handful that had been approved³. They found that the PDXs responded to approved drugs just as human responses predicted. And all the data Williams and her colleagues have collected since then suggest that tumours in PDXs respond as they do in people.

But when Golub and his colleagues reanalysed the data, they found three cases in which genome changes might have altered the outcome of the testing. Golub doesn’t think that PDX mice should work any better than mice implanted with cell lines. “I just don’t see the PDXs as being some magically different thing,” he says.

Golub and a colleague have argued for an international effort to establish more than 10,000 cancer cell lines⁴. This would be a boon, says David Weinstock, an oncologist at Harvard Medical School, and might obviate the need for PDX mice. But there are fewer than 2,000 cell lines available right now, and generating new ones is tricky. And although xenograft mice from these lines could be valuable, researchers have had more success in skipping the cell-line step to make PDX mice directly. “We’ve made 350 leukaemia and lymphoma models in one

laboratory with not that much money and not that much expertise,” Weinstock says. “We can’t make 350 cell lines.”

A MORE-HUMAN MOUSE

The real Achilles heel of PDX mice, however, is that to get the tumours to grow, researchers must use animals that lack an immune system. That makes it impossible to use PDXs to test immunotherapies. Several groups are now working to change that.

The Jackson Laboratory in Bar Harbor, Maine, takes stem cells from a human umbilical cord and injects them into mice that are a few weeks old. These stem cells differentiate and form some parts of the human immune system, mostly T cells. The researchers then transfer human tumours into those mice. “Nobody thought this would work,” says James Keck, a cancer researcher at the laboratory, because the umbilical-cord donor doesn’t match the tumour donor, so the T cells should attack the tumour. But the tumours have defence mechanisms to block the immune system, so “nine times out of ten, the tumour still grows,” says Keck. That has allowed scientists to test immunotherapies in a mouse model with human immune cells.

And just like in humans, these therapies don’t always work. For instance, Keck and colleagues have found that pembrolizumab, which ramps up the T-cell response, curbs bladder-cancer growth in mice carrying stem cells from one donor but not in mice carrying cells from another, even though both mouse types carried pieces of the same tumour⁵. “We’re actually getting close to what everybody has been asking for: a mouse model that mimics what’s going to happen in the clinic,” Keck says.

Ideally, researchers would like to create mice with tumour and immune cells from the same person. Meenhard Herlyn, who studies melanoma at the Wistar Institute in Philadelphia, Pennsylvania, and his colleagues are trying to use skin or blood cells from a patient to generate induced pluripotent stem cells, which could then be used to create immune cells. The model is almost complete, Herlyn says.

But even these next-generation PDX models have drawbacks. For example, human connective and vascular tissues in the tumour transplants are gradually replaced by mouse equivalents as they pass between mice.

Still, Keck is excited about the possibilities. “This is not your dad’s or mom’s xenograft any more,” he says. “These are models of complexity. We’ve now gone into a whole new level of oncology research.” ■

Cassandra Willyard is a freelance science journalist in Madison, Wisconsin.

1. Izumchenko, E. *et al.* *Ann. Oncol.* **28**, 2595–2605 (2017).
2. Ben-David, U. *et al.* *Nature Genet.* **49**, 1567–1575 (2017).
3. Gao, H. *et al.* *Nature Med.* **21**, 1318–1325 (2015).
4. Boehm, J. S. & Golub, T. R. *Nature Rev. Genet.* **16**, 373–374 (2015).
5. Wang, M. *et al.* *FASEB J.* **32**, 1537–1549 (2018).



Police in Seattle, Washington, wear masks to protect themselves during the 1918 flu pandemic that killed nearly 50 million people.

THE GHOST OF INFLUENZA PAST

A child's first flu infection shapes her response to all later ones. Now, researchers are realizing how important this 'imprint' is.

BY DECLAN BUTLER

By the time she is about three years old, a child has usually endured her first influenza infection. If it's a nasty bout, her temperature will rise and her muscles will ache. She's probably young enough that she won't recall the illness — but her immune system will.

When the virus enters her body, its presence prompts a pool of immature, unprogrammed immune cells to start competing to become the flu's tracker and assassin. The winners — cells that bind most strongly to the virus — store a memory of the pathogen, ready to recognize and attack it the next time it strikes.

But influenza is an inveterate shape-shifter. Regions of its outer proteins can mutate as it replicates, allowing it to avoid immune detection. When infections with new flu strains occur later in life, the immune system will mount a response based on that first

encounter, reacting strongly to recognized regions of the virus, but not to any that have changed. Immune cells can't tailor any novel antibodies that could help.

How exactly the immune system 'imprints' on its first-encountered strains presents a tantalizing puzzle to flu researchers — and solving it could help to combat the virus and improve vaccines.

Scientists suspect that understanding how imprinting works could help them to predict who will suffer most from seasonal strains and pandemics. Mounting evidence suggests that some people fare worse in deadly flu pandemics because their first childhood exposure was to a different version of the virus. Researchers think that this is why young adults experienced higher mortality than other age groups during the deadly 1918 pandemic¹, in which an estimated 50 million people died worldwide.

Knowledge of imprinting could help virologists to develop more-effective seasonal vaccines that could counteract circulating strains for several years, and a long-sought universal flu vaccine that could protect people for life against entirely new — and potentially pandemic-provoking — subtypes of flu. Imprinting seems to offer some immunity to flu strains related to the first infection. This broad immunity is often seen as a sign that the immune system could be coaxed into offering wide protection. “It does give us hope that we may be able to elicit a broadly protective immune response,” says Aubree Gordon, an epidemiologist at the University of Michigan in Ann Arbor.

Existing flu vaccines could certainly do with some help. Their effects wear off after a few months, and they aren’t very effective even in that brief window; during the 2017–18 flu season in the United States, people who received the vaccine were only 36% less likely to contract flu than those who weren’t immunized, although vaccination can lessen the severity of symptoms in those who do become ill.

Imprinting might help to explain these shortfalls. But right now, the mechanisms behind this process are poorly understood, says Jennifer Nayak, a paediatric immunologist at the University of Rochester Medical Center in New York. Getting to grips with imprinting will be important to researchers who hope to tailor a universal vaccine to fit people with different past flu exposures, says Scott Hensley, a viral immunologist at the University of Pennsylvania in Philadelphia. “The same vaccine given to different people will likely elicit different immune responses, depending on their history,” he says.

In April, the US National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, called for researchers to pitch projects that would explore the effects of imprinting on immunity, as part of a wider effort to fund research into a universal flu vaccine. The agency plans to spend US\$5 million on a large cohort study that will recruit and monitor infants from birth for at least three flu seasons to explore at the molecular level how their immune systems respond to initial exposure and subsequent flu infections and vaccinations. Immunizations are usually recommended for babies over 6 months of age.

Studying the virus can offer only so much; better protection will also depend on studying people. Researchers are realizing that the body can mount a surprisingly broad response, even against a shape-shifter like the flu. “Influenza is one of the best studied viruses on the planet,” says Katelyn Gostic, an epidemiologist at the University of California, Los Angeles (UCLA). “We’re discovering a whole new continent in a world that we thought was already well mapped.”

FLU FOUNDATIONS

The concept of imprinting was first proposed by the late Thomas Francis, a virologist and epidemiologist at the University of Michigan, whose studies in the 1940s and 1950s were the first to show that individuals generate stronger antibody responses to the first flu strain they encounter, compared with those they’re exposed to later in life².

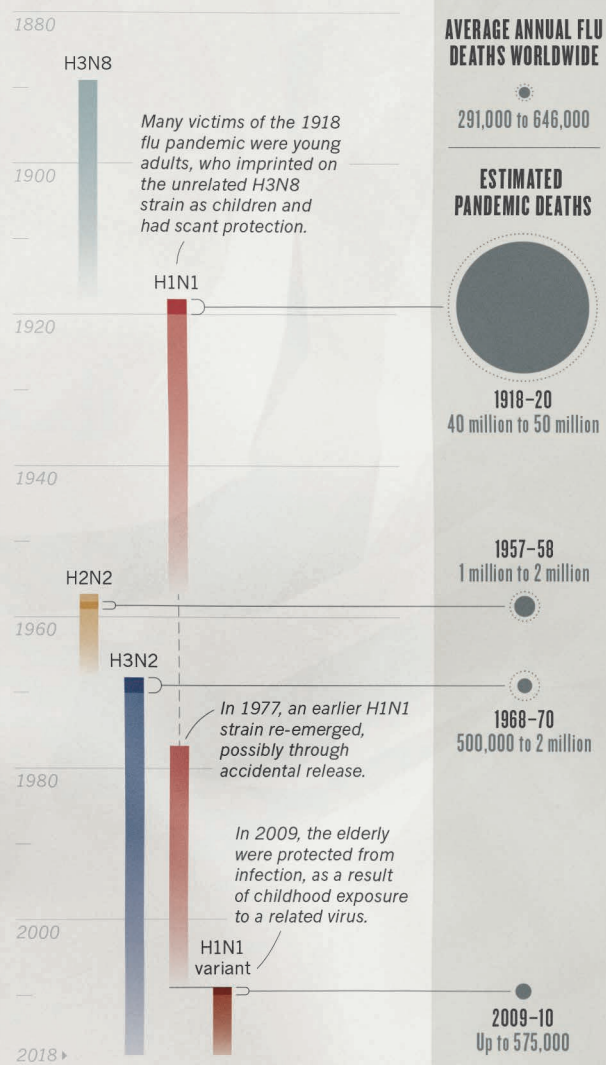
Researchers have since refined the concept. In a study of more than 150 people aged 7–81 in southern China, scientists measured antibody levels against several different strains of flu virus, looking at how their immune systems responded to strains they would have encountered at different times in their lives. The researchers found that after the first infection, subsequent strains have a progressively dwindling influence on the immune response³, explains Justin Lessler, an epidemiologist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, and a co-author of the study. “While immune imprinting plays a critical role, a focus on that alone can lead us to miss important aspects of how influenza immunity develops across multiple exposures,” he says.

In 2009, a new flu variety emerged in Mexico, resulting in a pandemic that gave researchers one of their best chances yet to study imprinting using modern immunological methods. A series of studies^{4,5} suggests that the virus prompted such a strong immune response that it ‘awoke’ in people who contracted it a broad immunity that had lain dormant since early imprinting. Many individuals generated antibodies that could attack not only the new strain but also members of its wider family.

Flu viruses come in a few flavours. The major version that causes illness

REIGNING STRAINS

Different subtypes of influenza have emerged over time, sometimes provoking pandemics. The subtypes circulating in the year you were born can influence your response to pandemic strains, strengthening your defence against similar versions, but making you more susceptible to different subtypes.



in humans has many subtypes, which are named after proteins on their surface: there are 18 known forms of the haemagglutinin (HA) protein, and 11 of the neuraminidase (NA) protein. Each virus subtype has an HA and an NA variant. Bolting them together gives each subtype its name — such as H1N1 or H3N2. Some have been found to infect only certain animal groups, but others can morph into new versions capable of infecting humans.

In a *Science* paper⁶ in 2016, Gostic and her colleagues analysed all known human cases of two subtypes of bird flu, H5N1 and H7N9, in circulation in six countries. The two viruses afflicted different age groups. H5N1 mostly infected young people, whereas almost all cases of H7N9 were in older people. By looking at the year of birth of each individual with the flu, they found that susceptibility abruptly changed in 1968, with people born before then more vulnerable to H7N9, and those born after more vulnerable to H5N1.

These people hadn’t met either subtype before. But depending on when they were born, they had encountered related varieties. Flu subtypes can be divided into two groups according to certain characteristics of their HA protein. H5N1 belongs to the same broad group as H1N1



Frozen flu-virus strains are stored at the US National Institutes of Health.

and H2N2 — strains that circulated seasonally before 1968.

Anyone born before that year would have been imprinted with one of these group 1 strains, and so protected from H5N1. But in 1968, everything changed: a pandemic of H3N2 struck, and became the only seasonal subtype. Most people born after that time were thus imprinted with the H3N2 strain, a group 2 virus. The H7N9 variant belongs to the same group — so many people born after 1968 were protected against it.

The finding suggests that imprinting with a virus from one of the two HA groups might offer broad cross-protection against new subtypes in the same group, challenging many public-health experts' assumption that most people would have little or no protection in pandemics, which are usually caused when new subtypes of flu emerge.

"The strength of the protective effect against severe H5N1 and H7N9 infection was shocking," says disease ecologist James Lloyd-Smith, co-author of the paper and also at UCLA. Using modelling, the researchers showed that childhood imprinting gave 75% protection against severe disease and 80% against death from these avian flu viruses.

Variations in susceptibility among different age groups have been observed in other pandemics. In the 1918 pandemic, perpetrated by an H1N1 subtype, those most severely affected were young adults with broad protection against H3N8, which circulated between 1889 and 1918 when they were children. H3N8 belongs to a different group from H1N1 (see 'Reigning strains'). The 2009 pandemic was caused by a variant of H1N1, but even so, there were very few cases in the elderly, who would have imprinted on the earlier version of H1N1 that circulated after the 1918 pandemic, says Patrick Wilson, an immunologist at the University of Chicago in Illinois. An H1N1 virus also appeared in the 1970s: it was so similar to a previous strain that scientists think it was accidentally released from a laboratory or a vaccine trial⁷. "It's sort of fun to look at when you were born and sort of infer what your first imprint was," Hensley says.

The priority now is to work out how the human body is imprinting on the first strains it sees. "We need to tease out what the immunological basis of that is," says Hensley.

Over the past decade, researchers have been building a palette of techniques for studying imprinting at the molecular level. It's easy to test the level of all antibodies generated in response to a bout of flu, for instance, but getting to the roots of imprinting requires being able to focus on the subsets of antibodies that generate broad immunity.

For example, researchers are now able to sort and analyse hundreds of thousands of single cells, and they can use single-cell sequencing to characterize the major players of the immune system before and after the cells respond to their first infection. Scientists would like to know how those cells engineer such a long-lasting response to future flu.

"Now, our tools are much more refined, providing an extremely granular look at what is occurring upon first exposure, and repeated exposure, to influenza and influenza vaccine," says Buddy Creech, director of the

Vanderbilt Vaccine Research Program at Vanderbilt University Medical Center in Nashville, Tennessee. He is co-directing the Universal Influenza Vaccine Initiative, a multi-university project that launched last October to study the immune response to flu and how broad immunity might be evoked. Once those mechanisms are better understood, they might be recapitulated to help make vaccines more broadly active, says Nayak.

PEOPLE POWER

For researchers wishing to apply these tools, funders such as the US National Institutes of Health and the Bill & Melinda Gates Foundation are stepping in to help.

The Gates Foundation announced a \$12-million tranche of funding in April, which it plans to put towards pilot projects that aim to develop universal flu vaccines; the call mentions imprinting and other features of the host's immune response, and will prioritize higher-risk ventures.

In the same month, the NIAID issued its \$5-million call for proposals to follow large numbers of children over at least three flu seasons and potentially for years afterwards. The ultimate goal of the study, according to the NIAID, is to provide information that will help researchers to design long-lasting, universal vaccines.

Until now, research into childhood exposure has been limited, so the NIAID call is welcome news, says Nayak. Most studies of flu in children have been small, and haven't characterized each individual's exposure history firmly enough, she says. "This makes it impossible to even address whether imprinting is occurring, much less determine the mechanism responsible."

Part of the problem has been in tracking an infant's immune system, which requires repeated blood draws. As recently as 5 years ago, assays required drawing 10–20 millilitres of blood, making immunological monitoring of young babies impractical (a 3-kilogram newborn has only 240 millilitres of blood). But advances in technology have overcome that obstacle. "With these single-cell assays, you can do strong immunological work-ups with just 1 to 2 millilitres of blood," says Hensley, who has applied to run a study using cohorts in the United States and Hong Kong.

These techniques will enable researchers to catalogue an infant's exposures and vaccinations precisely over time, and to sketch a detailed picture of how immunity differs when stimulated by natural infection compared with vaccination.

The NIAID call aims to complement other cohort studies of flu around the world. The agency already supports influenza research through cohorts in Nicaragua, Hong Kong and New Zealand, but none focuses on childhood imprinting. Gordon runs the Nicaraguan cohort, which is studying the incidence and severity of flu in children. Hers is the only large cohort set up to enrol and follow children from birth, and so is well placed to study imprinting. She has applied for NIAID funding as part of a consortium, to enable her team to incorporate the specialized immunology expertise needed.

Nayak already has a small pilot imprinting study in progress, which has so far enrolled 129 children since it started in late 2016. She, too, has put in a bid to the NIAID, involving the University of Rochester and the University of Minnesota, with cohort sites in the United States and Australia. Having multiple sites hedges against the risk of a few quiet flu seasons, or a few seasons dominated by just one flu flavour.

For scientists who want to chase the elusive universal flu vaccine, the cohort studies are one strand of a multi-pronged strategy. They will also need to research basic viral biology and find fresh ingredients for vaccines, says Creech. "We really have to work the problem from both sides." ■

Declan Butler is a senior reporter for Nature.

1. Gagnon, A., Acosta, J. E., Madrenas, J. & Miller, M. S. *PLoS Pathog.* **11**, e1004615 (2015).
2. Francis, T. Jr *Proc. Am. Philos. Soc.* **104**, 572–578 (1960).
3. Lessler, J. et al. *PLoS Pathog.* **8**, e1002802 (2012).
4. Wrammert, J. et al. *J. Exp. Med.* **208**, 181–193 (2011).
5. Guthmiller, J. J. & Wilson, P. C. *Curr. Opin. Immunol.* **53**, 187–195 (2018).
6. Gostic, K. M., Ambrose, M., Worobey, M. & Lloyd-Smith, J. O. *Science* **354**, 722–726 (2016).
7. Rozo, M. & Gronvall, G. K. *mBio* **6**, e01013-15 (2015).

COMMENT

COMMUNITY Jordanian biologist's take on science, ethics and life **p.164**



PHYSICS Refracting quantum physics through the double-slit experiment **p.165**

NEUROSCIENCE Exploration of genetic factors behind mental disorders **p.166**

CONSERVATION Monitor impacts of tourists on Pacific reefs **p.167**

CHINA MINISTRY OF TRANSPORT/AP/REX/SHUTTERSTOCK



The *Sanchi* oil tanker on fire in the East China Sea after a collision with a cargo vessel in January 2018.

Human errors are behind most oil-tanker spills

Misleading accident data sets skew research and laws.
Zheng Wan and Jihong Chen set out three priorities.

In January, the oil tanker *Sanchi* collided with a cargo vessel in the East China Sea, 300 kilometres off Shanghai, China. The tanker caught fire, exploded and sank, killing all 32 members of its crew and spilling or burning more than 100,000 tonnes of petroleum products. In May, China's Maritime Safety Administration gave its final verdict: both vessels had violated navigational protocols and watch-keeping codes¹. Although accidents such as this are now rare, we fear that they could be set to increase.

Assuming much of its cargo entered the sea, *Sanchi* could be one of the largest such spills in nearly 30 years, since the *Exxon Valdez* dumped 37,000 tonnes of crude oil into Alaska's Prince William Sound in 1989. Even as the quantity of oil and gas transported by sea has doubled since the 1970s, there have been fewer spills greater than 7 tonnes — down from roughly 80 per year to about 7 per year (see 'Tanker trends'). Double hulls and fire-fighting systems that use inert gases have helped.

Two trends in the past decade threaten those improvements. First, the accident rate for major tankers (those that carry more than 15,000 tonnes, with and without spills) almost tripled between 2008 and 2017: from 1 accident for every 40 tankers to 1 in every 15 (ref. 2). Second, to cut costs, substandard ships with poor maintenance records and unqualified personnel are increasingly registered in countries that have lax regulation. The chance of a major spill occurring in a region that is unable to cope could rise, putting fragile coasts at risk.

Any spill is disastrous — ecologically, economically and socially. The *Exxon Valdez* disaster killed an estimated 250,000 seabirds, hundreds of otters, seals and eagles, and some two dozen killer whales. Oil vapours are toxic and contaminate seafood, harming public health and the local economy. And residues linger for decades³. Large spills, such as from the *Tasman Spirit*, which ran aground off Karachi in 2003, or from the *Prestige* that sank in 2002 off Galicia, Spain, cause billions of dollars in damages⁴. Clean-ups can cost more than US\$20,000 per tonne of oil spilt⁵.

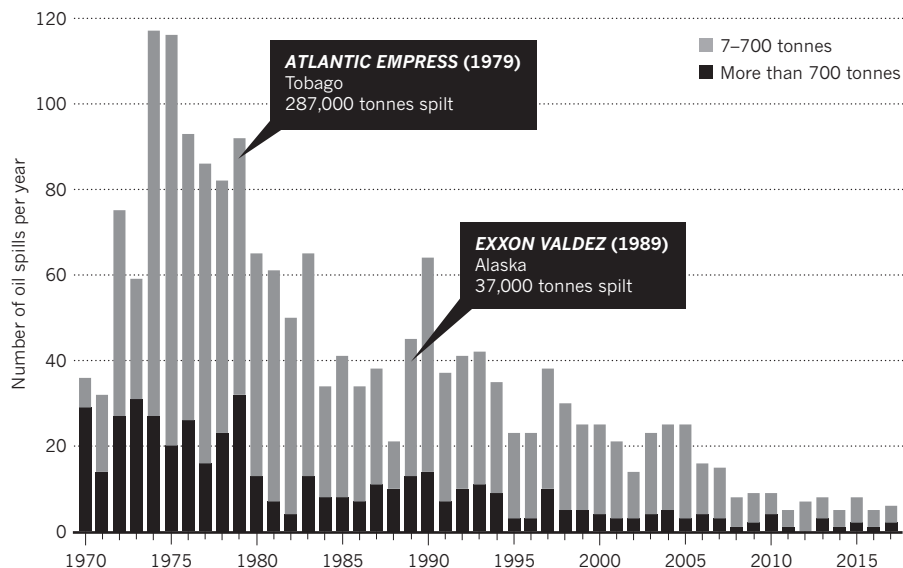
For the *Sanchi* spill, the ecological impacts, legal implications and clean-up strategies are unclear. The tanker was carrying 136,000 tonnes of condensate oil, a volatile and toxic hydrocarbon compound that is generated during the processing of natural

TANKER TRENDS

The number of large oil spills from tankers has fallen since 1970. But increases in the volume of oil traded around the world and in accident rates could lead to more spills in the future.

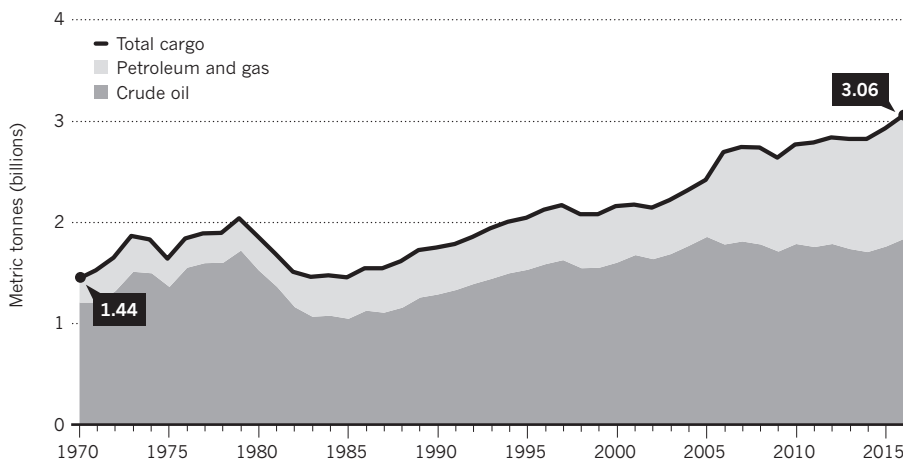
FEWER SPILLS

Double hulls and fire systems have reduced tank breaches.



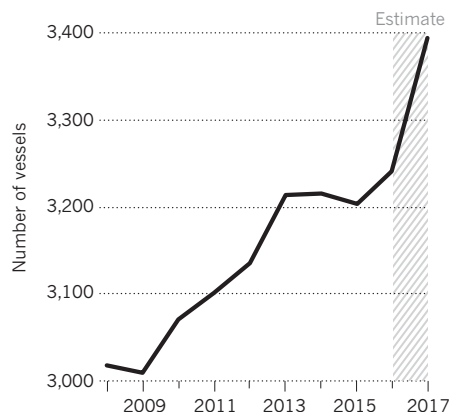
MORE CARGO

Rising energy use is increasing the volume of oil transported.



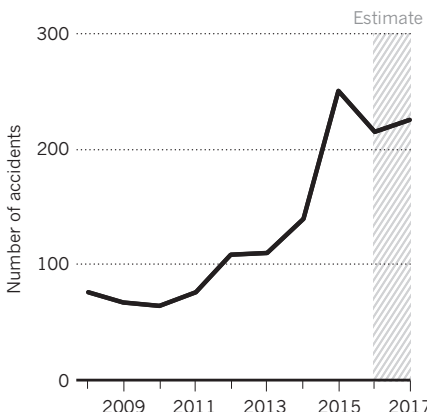
MORE SHIPS

A growing number of registered oil tankers have a carrying capacity of more than 15,000 tonnes.



MORE MISTAKES

Accidents are rising as trade grows and shipping lanes become congested.



▶ gas. Much of the unburnt cargo was spilt, together with 2,000 tonnes of the ship's own fuel. Ecological damage is inevitable. The area is a spawning ground for fish such as bluefin leatherjacket (*Thamnaconus septentrionalis*) and largehead hairtail (*Trichiurus japonicus*), and invertebrates such as swordtip squid (*Uroteuthis edulis*). It lies on a migratory route for at least three species of whale (see go.nature.com/2msmwn9). But no nation is duty-bound to issue an ecological assessment, because the accident happened in the high seas, beyond local jurisdictions. Neighbouring nations such as Japan and South Korea are keeping an eye on the situation. The tanker was Iranian-owned and registered in Panama. The cargo ship was based in Hong Kong.

Reducing accidents is the obvious answer. But the causes are widely misunderstood. Shipping records often list consequences — collisions, groundings and explosions — rather than reasons, such as poor navigation, lack of maintenance, miscommunication and other human errors. Researchers studying these databases thus reach the wrong conclusions and propose inappropriate policies. Tighter regulations on how ships are built do nothing if they go unenforced.

Clean-up technologies also need to improve to minimize damages from spills. The oil and shipping industries still use decades-old techniques, such as mixing chemical dispersants with oil-contaminated seawater. The dispersants break up the slicks into droplets that should, in theory, be easier for microorganisms to break down. But reactions can make the combination toxic to species such as rotifers (zooplankton at the base of the marine food web)⁶. There have been few long-term environmental studies of the ecological impacts of dispersants.

Researchers need to refocus discussions about tanker safety on to the human behaviours that cause accidents and on to improving safety protocols. They need to re-evaluate the risk-prediction models that are used to judge how often to inspect ships. Clean-up technologies also need to be improved and commercialized.

MULTIPLE FACTORS

The world's 7,000 oil tankers comprise 14% of the total shipping fleet. Tankers accounted for three-quarters of oil spills larger than 200 tonnes between 1974 and 2010, or more than 60% of the 9.8 million tonnes of oil spilt during that time⁷. The remainder came from pipelines, exploration and production and refineries.

The blame stretches beyond just the shipping operators. Half of oil tankers are registered in nations that do little to oversee vessel safety and crew training. A dozen countries, notably Panama, Liberia, the Marshall Islands, the Bahamas and Malta, allow almost any ship to fly their 'flags of convenience'. Panama and Liberia — the

SOURCES: 'FEWER SPILLS', 'TOPF'; 'MORE CARGO', 'UNCTAD'; 'MORE SHIPS' & 'MORE MISTAKES', REF. 2

nations with the biggest fleets — control 18% and 12% of the world's shipping tonnage (8,000 and 3,000 ships, respectively). Between 1967 and 2017, 12 of the tankers involved in the top 20 spills flew a convenient flag; 9 of those were from Liberia.

Tighter regulations are set out in the 1986 United Nations Convention on Conditions for Registration of Ships. It has yet to enter into force owing to industry lobbying. At least 40 states with more than 25% of the world's shipping tonnage must sign on; only 14 have done so.

In the meantime, coastal nations inspect foreign-registered vessels that enter their ports to ensure they comply with international maritime conventions. Port authorities use predictive models of risks to decide which vessels to inspect and how often. For example, a 20-year-old ship carrying dangerous cargo with poor safety records might be checked every 6 months; a new ship with good safety records every 36 months. But parameters such as ship age or historical safety records are unreliable indicators of risk. Older vessels are often safer — they have survived owing to better-quality or well-maintained equipment⁸. And historical safety records can be subjective and misleading. The results are shaped by who inspected the ship and how.

Checks are no deterrent⁹. Tighter inspections with heavier penalties in tightly regulated countries merely shift shabby ships to less-regulated nations. There are few civil or criminal penalties for flouting rules. Ships can travel thousands of kilometres between checks.

Port inspections are costly, for both authorities with limited staff and shipping companies with tight operating schedules. The scope is limited. It is easier to check the completeness of documents, such as records of crew rest hours, than the integrity of information. Flawless records can indicate that the crew is aware of the safety standards, or that they know how to fool the system.

ACKNOWLEDGE CAUSES

Human errors are behind at least 80% of tanker accidents (see go.nature.com/2nwgubp). Such errors include fatigue caused by overwork, inadequate expertise on a specific operation, poor communication or the use of outdated navigational charts. Yet these are rarely listed as causes in databases of shipping accidents¹⁰. Such confusion thwarts research and risk management.

For example, in 1994, the *Nassia* tanker spilt around 13,500 tonnes of crude oil in Turkey's Bosphorus waterway. Records report that the tanker collided with another vessel, grounded and exploded. But other factors were not noted. For instance, the other vessel lost power and was unable to steer away from *Nassia*. The reason has not



Oil spills threaten seabirds and other marine life.

been established in this case, but inadequate maintenance and repairs are often a cause of engine problems (go.nature.com/2nymjsv).

Researchers often misinterpret the statistical results generated by oversimplified and improper classification data sets. Collisions, groundings and explosions, for example, are described as primary causes for tanker incidents even though they are consequences (see go.nature.com/2jaekte). There is little or no information about the crew and their employer. Routinely calling to limit these physical factors without understanding the real drivers creates an unrealistic sense of hope in the shipping community that advancing technologies can solve all the problems. Weak policy prescriptions follow, such as mandating that ships are resistant to grounding.

Policies would be more effective if they acknowledged the role of human error. For example, crew fatigue caused by long working hours and isolation is a significant contributor. Raising the minimum number of qualified crew can reduce average workload and help to prevent mistakes.

THREE PRIORITIES

Research on the following would limit risk and damage.

Improve port inspections. Researchers should re-evaluate the algorithms that are used to decide which ships are inspected and when. The local maritime authorities should conduct randomized and controlled trials to optimize inspection strategies. They can borrow experiences from predictive policing schemes that use machine learning in some cities to fight crime¹¹. Developed nations should provide aid for developing nations to ensure uniform standards.

Inspectors should look beyond records and, for example, conduct random interviews with crew members to judge whether they understand the safety protocols. We recommend they include surprise questions

to test how tanker crew members will react in a crisis.

Study human errors. The International Maritime Organization (IMO) should collaborate with the research community to better understand how human error contributes to shipping accidents. Accurate data sets that document the objective causes of the incidents are key. Researchers need to revisit previous oil-spill incidents and reclassify the causes. Types of human error can be identified through the investigation reports on the IMO's website.

The tanker industry must use these data to design better strategies for reducing human errors. For example, language is a barrier for many multicultural crews and could be improved through training.

Develop sustainable clean-up technologies.

New physical and mechanical clean-up methods should be developed. Promising methods are emerging, such as soak-up sponges, bioremediation and devices for separating oil and water. They still need to be commercialized. Chemists and toxicologists should evaluate chemical dispersants for efficacy and toxicity. Government agencies and the oil industry should prioritize the funding of such interdisciplinary research.

As our understanding improves, regulatory instruments must also evolve. States need to take responsibility for their fleets. For example, only vessels owned by capital invested by a certain country or that sail in its waters for a considerable time should be entitled to that nation's flag. Countries would thus have more incentive and be more able to exercise jurisdiction and control. The IMO should require that the tanker industry signs up to this reformed registration system first. As global energy demands grow, tanker safety must remain a priority. ■

Zheng Wan is associate professor and **Jihong Chen** is professor at the College of Transport and Communications, Shanghai Maritime University, China.
e-mail: jihongchen@vip.sina.com

1. China Maritime Safety Administration. *Report on the Investigation of the Collision between M.T. Sanchi and M.V. CF Crystal in East China Sea on 6 January 2018* (China MSA, 2018).
2. Vidmar, P. & Perković, M. *Safety Sci.* **105**, 178–191 (2018).
3. Peterson, C. H. *et al. Science* **302**, 2082–2086 (2003).
4. Alló, M. & Loureiro, M. L. *Ecol. Econ.* **86**, 167–175 (2013).
5. Kontovas, C. A., Psaraftis, H. N. & Ventikos, N. P. *Mar. Pollut. Bull.* **60**, 1455–1466 (2010).
6. Rico-Martinez, R., Snell, T. W. & Shearer, T. L. *Environ. Pollut.* **173**, 5–10 (2013).
7. Eckle, P., Burgherr, P. & Michaux, E. *Environ. Sci. Technol.* **46**, 13002–13008 (2012).
8. Li, K. X., Yin, J. & Fan, L. *Transp. Res. A: Policy Pract.* **66**, 75–87 (2014).
9. Knudsen, O. F. & Hassler, B. *Mar. Policy* **35**, 201–207 (2011).
10. Chen, J. *et al. J. Clean. Prod.* **180**, 1–10 (2018).
11. Shapiro, A. *Nature* **541**, 458–460 (2017).



Molecular biologist Rana Dajani at a workshop on educating refugees in 2015.

COMMUNITY

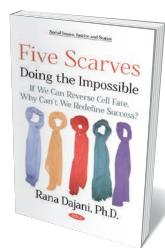
The real wins for women in science

Malak Abedalthagafi extols a memoir from a Jordanian biologist and trailblazer in women's rights.

Can a breakthrough in stem-cell research revolutionize feminism? Can a scientist apply the scientific method to her own life to find solutions to social problems? In *Five Scarves*, Jordanian molecular biologist Rana Dajani reveals with passion and cogency how she has explored those possibilities. She speaks to humanity's capacity to overcome challenges — not least, improving the treatment of women and children.

The book is part call to action, part research journal and part autobiography: the five scarves are the different 'hats' Dajani wears as scientist, mother, teacher, social entrepreneur and trailblazing feminist. She has long written and spoken about the obstacles facing women in academia, and how they vary by discipline and culture. As she notes, across the Middle East, women constitute just under 40% of researchers in science, technology, engineering and medicine; in the United States, a mere 24%. Moreover, as a champion of women's central role in families, she is determined to change mindsets so that — as she asserts — women worldwide do not have to choose between career and family. Having worked in both the United States and Saudi Arabia, I find that resonates with me.

Describing a 1970s childhood and adolescence in Jordan and the United States,



Five Scarves: Doing the Impossible — If We Can Reverse Cell Fate, Why Can't We Redefine Success?

RAANA DAJANI
Nova (2018)

in communities most affected by power struggles beyond their control. She stressed the importance of education, for instance, so that vulnerable people, especially children, are no longer mistreated or manipulated. And her extensive reading offered glimpses of far-flung travel and other opportunities.

Marrying in the early 1990s, she began a family while still in education; motherhood brought a determination to wear multiple scarves with grace. Dajani records that for her, pregnancy and birth were a revelation of

Dajani writes that she learnt from strong women how to be responsible for community well-being. Her mother taught her that in Islam, a person is judged on intention, and that every effort counts, however seemingly inconsequential. With parents from Syria and the Palestinian territories, Dajani became passionately outspoken on the rights of women and families, particularly

the profundity of human biology. She and her young family moved to Iowa City in 2000 so that she could pursue her PhD at the University of Iowa. Her husband gave up a career for the move; Dajani is optimistic that more men are supporting their wives in this way. Removing sexist assumptions and roles from family life is part of her redefinition of success.

She criticizes some would-be support in the United States. A number of tech giants offer to freeze employees' eggs to let them have children later. Yet the technology is not foolproof: a study by the UK Human Fertilisation and Embryology Authority found that in 2016, only 19% of implantation cycles using frozen eggs succeeded. Paid parental leave and childcare would be more just, pragmatic and economical, she argues. The United States is the only industrialized nation with no mandate for paid maternity leave.

In 2005, Dajani and her family returned to Jordan. At the Hashemite University in Amman, she researched the genetics of the country's Circassian and Chechen ethnic groups and began to collaborate with scientists worldwide, for instance on the study of ancient human lineages. In 2008, inspired by stem-cell breakthroughs, Dajani formed a committee on the political and ethical aspects of the research. That led to Jordan's Stem Cell Research and Therapy Law, which encouraged the work but regulated and decommercialized it, setting a precedent in the region.

From 2015, Dajani was involved in studies that helped participants to be part of their own success. One, which she spearheaded in Jordan, was initiated by medical anthropologist Catherine Panter-Brick to gauge the impact of a programme to reduce trauma in young Syrian refugees. One of Dajani's contributions was to explain the link between stress and levels of the hormone cortisol in hair; crucially, she and her team also ensured that the young people had agency, collecting their own data and helping to find new approaches.

Treating social challenges such as poverty and illiteracy as a science experiment, Dajani initiated the We Love Reading project in Jordan, hypothesizing that getting children excited about books would stir social change beyond their communities. Within 12 years, the programme had distributed 250,000 books and established 1,500 neighbourhood libraries. There is much more in this memoir, from Dajani's work setting up mentoring networks for female scientists in the Middle East to her bold, innovative approach to teaching.

In a sense, she asks: if molecules can communicate effectively, why can't we? ■

Malak Abedalthagafi is a physician-scientist, chair of genomics research at King Fahad Medical City and assistant research professor in molecular genetics and neuropathology King Abdulaziz City for Science and Technology, both in Riyadh, Saudi Arabia. e-mail: malthagafi@kacst.edu.sa

QUANTUM PHYSICS

Two slits, one hell of a quantum conundrum

Philip Ball lauds a study of a famous experiment and the insights it offers into a thoroughly maddening theory.

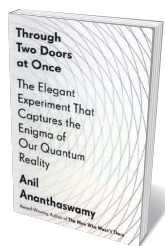
According to the eminent physicist Richard Feynman, the quantum double-slit experiment puts us “up against the paradoxes and mysteries and peculiarities of nature”. By Feynman’s logic, if we could understand what is going on in this deceptively simple experiment, we would penetrate to the heart of quantum theory — and perhaps all its puzzles would dissolve.

That’s the premise of *Through Two Doors at Once*. Science writer Anil Ananthaswamy focuses on this single experiment, which has taken many forms since quantum mechanics debuted in the early twentieth century with the work of Max Planck, Albert Einstein, Niels Bohr and others. In some versions, nature seems magically to discern our intentions before we enact them — or perhaps retroactively to alter the past. In others, the outcome seems dependent on what we know, not what we do. In yet others, we can deduce something about a system without looking at it. All in all, the double-slit experiment seems, to borrow from Feynman again, “screwed”.

The original experiment, as Ananthaswamy notes, was classical, conducted by British polymath Thomas Young in the early 1800s to show that light is a wave. He passed light through two closely spaced parallel slits in a screen, and on the far side saw several bright bands. This, he realized, was an ‘interference’ pattern. Caused by the interaction of waves emanating from the openings, it’s not unlike the pattern that appears when two pebbles are dropped into water and the ripples they create add to or dampen each other’s peaks and troughs. With ordinary particles, the slits would act more like stencils for sprayed paint, creating two defined bands.

We now know that quantum particles create such an interference pattern, too — evidence that they have a wave-like nature. Postulated in 1924 by French physicist Louis de Broglie, this idea was verified for electrons a few years later by US physicists Clinton Davisson and Lester Germer. Even large molecules such as buckminsterfullerene — made of 60 carbon atoms — will behave in this way.

You can get used to that. What’s odd is that the interference pattern remains — accumulating over many particle impacts — even if particles go through the slits one at a time. The particles seem to interfere with themselves. Odder, the pattern vanishes if we use



Through Two Doors at Once: The Elegant Experiment That Captures the Enigma of Our Quantum Reality
ANIL ANANTHASWAMY
Dutton (2018)

a detector to measure which slit the particle goes through: it’s truly particle-like, with no more waviness. Oddest of all, that remains true if we delay the measurement until after the particle has traversed the slits (but before it hits the screen). And if we make the measurement but then delete the result without looking at it, interference returns.

It’s not the physical act of measurement that seems to make the difference, but the “act of noticing”, as physicist Carl von Weizsäcker (who worked closely with quantum pioneer Werner Heisenberg) put it in 1941. Ananthaswamy explains that this is what is so strange about quantum mechanics: it can seem impossible to eliminate a decisive role for our conscious intervention in the outcome of experiments. That fact drove physicist Eugene Wigner to suppose at one point that the mind itself causes the ‘collapse’ that turns a wave into a particle.

Ananthaswamy offers some of the most lucid explanations I’ve seen of other interpretations. Bohr’s answer was that quantum

mechanics doesn’t let us say anything about the particle’s ‘path’ — one slit or two — before it is measured. The role of the theory, said Bohr, is to furnish predictions of measurement outcomes; in that regard, it has never been found to fail. (However, he did not, as is often implied, deny that there is any physical reality beyond measurement.) Yet this does feel rather unsatisfactory. Ananthaswamy seems tempted by the alternative idea offered by David Bohm in the 1950s. Here, quantum objects are both particle and wave, the wave somehow ‘piloting’ the particle through space while being sensitive to influences beyond the particle’s location. But Ananthaswamy concludes that “physics has yet to complete its passage through the double-slit experiment. The case remains unsolved.”

With apologies to researchers convinced that they have the answer, this is true: there is no consensus. At any rate, Bohr was right to advise caution in how we use language. There is nothing in quantum mechanics as it stands, shorn of interpretation, that lets us speak of particles becoming waves or taking two paths at once. And there is no reason to regard the wavefunction as more or less than an abstraction. This mathematical function, which embodies all we can know about a quantum object (and features in the iconic equation devised by Erwin Schrödinger to describe the object’s wave-like behaviour) was characterized rather nicely by physicist Roland Omnès. He called it “the fuel of a machine that manufactures probabilities” — that is, probabilities of measurement outcomes.

Refracting all of quantum mechanics through the double slits is both a strength and a weakness of *Through Two Doors at Once*. It brings unity to a knotty subject, but downplays some important strands. Those include John Bell’s 1964 thought experiment on the nature of quantum entanglement (conducted for real many times since the 1970s); the role of decoherence in the emergence of classical physics from quantum phenomena (adduced in the 1970s and 1980s); and the emphasis on information and causality in the past two decades. Still, given that popularization of quantum mechanics seems to be the flavour of the month — summoning Adam Becker’s 2018 book *What is Real?*, Jean Bricmont’s 2017 *Quantum Sense and Nonsense*, a forthcoming book by physicist Sean Carroll, and my own 2018 *Beyond Weird* — there’s no lack of a wider perspective.

And we need that. Ananthaswamy’s conclusion — that perhaps all the major interpretations are “touching the truth in their own way” — is not a shrugging capitulation. It’s a well-advised commitment to pluralism, shared with Becker’s book and mine. For now, uncertainty seems the wisest position in the quantum world. ■

Philip Ball is a writer based in London.
e-mail: p.ball@btinternet.com



Bands of light in the double-slit experiment.

Biopsychiatry and the mind

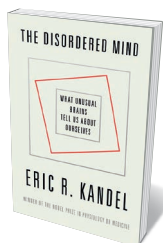
Douwe Draaisma weighs up Eric Kandel's study on mental illnesses as brain diseases.

In 1939, nine-year-old Eric Kandel and his family fled Nazi-occupied Vienna for the United States. Eventually entering Harvard University in Cambridge, Massachusetts, he intended to train as a psychoanalyst. But he felt that understanding mechanisms such as repression, conceived by fellow Viennese exile Sigmund Freud, demanded knowledge of their neurological underpinnings. So he turned to brain research.

Kandel's work in the 1960s — uncovering the circuitry of learning processes by measuring neuronal activity in the marine mollusc *Aplysia* — earned him a Nobel prize. But his first intellectual passion never left him. Over the past 15 years or so, he has attempted to restore prestige and influence to psychoanalysis by wedding it to a “new biology of mind”. This, he argues in *The Disordered Mind*, is aided by three advances: brain imaging, study of how psychiatric disorders are inherited and animal models for conditions such as autism spectrum disorder (ASD).

Like Freud, he probes the unconscious. But for Kandel, these are processes of genetic development and dysfunctions of neurological circuits. Clearly and concisely, he leads us through recent findings and hypotheses on various disorders. Some are neurological, such as Alzheimer's, Huntington's and Parkinson's diseases. Some have been interpreted as psychiatric, such as depression, bipolar disorder and schizophrenia. The modern perspective, he asserts, is that these are ultimately brain disorders. ‘Modern’ might be a misnomer: as long ago as 1845, German psychiatrist Wilhelm Griesinger issued the dictum *Geisteskrankheiten sind Gehirnkrankheiten* (mental illnesses are brain diseases). But Kandel is right about the importance of new tools — the methods, instruments and theories now available that might open promising avenues into psychiatric disorders.

The genetics of ASD and schizophrenia are a case in point. Autism is often diagnosed when a child is three or four. Schizophrenia is mostly late-onset, developing in adolescence or early adulthood. When one identical twin has autism, the chances are about 90% that the other has, too; with



The Disordered Mind: What Unusual Brains Tell Us About Ourselves
ERIC R. KANDEL
Farrar, Straus and Giroux (2018)



Brain imaging, including magnetic resonance scans, can provide information on psychiatric disorders.

schizophrenia, the figure is about 50%, even for twins raised apart. However, genetic studies have pointed out that a single mutation on chromosome 7 increases the chances of developing autism and schizophrenia. Other, spontaneous, types of mutation are more frequent in the sperm of older fathers, again raising the chances of developing ASD and schizophrenia. At the genetic level, there might be a close relationship between the conditions that fails to manifest in their expression.

Kandel notes, however, that in terms of brain development, ASD and schizophrenia are polarized. From the age of two, children with autism have been found to have more synapses (the contacts between neurons) than neurotypical children. Thus, autism might be correlated with insufficient ‘pruning’ of disused connections; an excess of synapses could explain the extraordinary long-term memory of some people with ASD. By contrast, people with schizophrenia have fewer synapses than do those without, raising the possibility that excessive pruning could underlie the disorder. A shortage of connections in the prefrontal cortex could cause problems with working memory and higher cognitive functions.

Kandel explores another case in which a common genetic change might cause very different symptoms: Alzheimer's and Parkinson's disease. Mutations associated with Alzheimer's might result in clumps of misfolded proteins in the brain, causing severe memory loss. Those associated with Parkinson's could lead to the death of cells that make dopamine, the neurotransmitter vital for movement.

None of these insights has brought relief,

pharmacological or otherwise. A potentially beneficial discovery, Kandel shows, is that healthy bones make osteocalcin, a hormone that stimulates proteins needed for memory formation. Studies by him and his colleagues show that age-related memory decline can be reversed when old mice are injected with osteocalcin. In humans, exercise builds bone mass, which could improve memory. (To misquote US President John F. Kennedy: “Ask not what your memory can do for you; ask what you can do for your memory.”)

Kandel's attempt to biologize psychiatry is not for the sensitive; his focus is medication and compensating for faulty wiring, not gaining psychological insight into inner turbulence. At times, he proposes a less-than-convincing reframing: that because psychoanalysis is a learning process, which involves synaptic changes, the therapy is essentially a biological treatment. However, reading a book or watching a film will bring about synaptic changes, too — and we wouldn't count either as primarily biological activities.

Bold propositions such as Kandel's in *The Disordered Mind* blur the distinction between therapies involving medication or surgery and those that use behavioural and cognitive means. Still, one should appreciate Kandel's humanistic aims: knowing more about disorders makes us less likely to stigmatize those who think or act differently. ■

Douwe Draaisma is professor of the history of psychology at the University of Groningen in the Netherlands.
e-mail: d.draaisma@rug.nl

Correspondence

Identify and punish ozone depleters

Emission rates of the ozone-depleting chlorofluorocarbon CFC-11 are no longer in decline (see S. A. Montzka *et al.* *Nature* **557**, 413–417; 2018). We suggest that Asia's construction boom could be part of the cause, by provoking a rapid increase in the unauthorized production of this chemical for building-insulation materials.

The 1987 Montreal Protocol resulted in a global ban on production of CFC-11. However, production has resumed since 2013 in some parts of China (see go.nature.com/2mj8ijg), coincident with the country's increased demand for insulation foam (C. Yang *et al.* *Energy Build.* **87**, 25–36; 2015). The same could be happening elsewhere, so other offenders urgently need to be identified.

If the fragile stratospheric ozone layer is to recover, the production and disposal of building-insulation materials must be more effectively monitored and managed, backed by stricter legislation. The development of low-cost alternatives to ozone-depleting substances for building materials is also a priority, given the pace of urbanization in China and other nations.

Hong Yang *University of Reading, UK.*

Roger J. Flower, Julian R. Thompson *University College London, UK.*
hongyanghy@gmail.com

Glacier engineering must mind the law

Proposals such as those of John Moore and colleagues (*Nature* **555**, 303–305; 2018) for Antarctic glacier geoengineering understate the legal challenges presented by the Antarctic Treaty System (ATS). This system is crucial to Antarctic governance, but faces considerable geopolitical pressure (*Nature* **558**, 161; 2018). It is essential that any activities

affecting the Antarctic ecosystem properly engage with the ATS from the outset.

Antarctic geoengineering proposals would not “require global consent” as Moore *et al.* state, but instead would need the approval of the 29 consultative parties to the 1959 Antarctic Treaty. The Scientific Committee on Antarctic Research is an important independent contributor to the ATS. However, it is actually the Committee for Environmental Protection (CEP), created by the 1991 Madrid Protocol to the Antarctic Treaty, that formally advises the consultative parties about proposals affecting the Antarctic environment.

The Madrid Protocol bans mining and declares Antarctica a natural reserve. We think that the CEP is likely to advise that the “major disturbances to local ecosystems” arising from Moore and colleagues' proposals — particularly quarrying of local rock and dredging — would infringe Madrid Protocol protections. Geoengineering that affects marine ecosystems might also require separate permission under the 1982 ATS Convention on the Conservation of Antarctic Marine Living Resources.

Any discussion of geoengineering in Antarctica needs to preserve and strengthen Antarctic governance, not weaken it. This is a task for international lawyers and policymakers as well as scientists.

Brendan Gogarty* *University of Tasmania, Hobart, Australia.*
brendan.gogarty@utas.edu.au

*On behalf of 6 correspondents (see go.nature.com/2kjaady for full list).

Underpin tourism regulation with data

We understand the concerns of Philippe Borsa and colleagues over the New Caledonia government's plans to open the Chesterfield reefs to ecotourism cruise ships (*Nature* **558**, 372; 2018). In our view as conservation

biologists, conservationists also need to consider context — such as the benefits that tourism could bring to the islands' fragile economy — and to discuss with government how to make such tourism sustainable.

The Natural Park of the Coral Sea, which harbours the reefs, is one of the largest marine protected areas in the world. As it becomes increasingly autonomous, New Caledonia is legitimately looking for ways to diversify its economy and is turning to the resources offered by its maritime exclusive economic zone. The zone measures 1.74 million square kilometres and hosted 219 cruise liners carrying some 500,000 passengers in 2017.

We call on the scientific community to work with local authorities in guiding New Caledonia towards sustainable use of its wild and remote oceanic space. More data are needed on the seabirds that inhabit the fragile, low-lying island ecosystems in these areas, including on the ecological and behavioural consequences of human incursions on seabird breeding. These data must be shared openly. Lessons can also be learnt from tourism management of other tropical islands, such as those associated with Australia's Great Barrier Reef, and in the polar regions.

Eric Vidal, Martin Thibault, Karen Bourgeois *IRD Research Center, Nouméa, New Caledonia.*
eric.vidal@ird.fr

Political pressures on Romania's research

Last month's Rectors' conference of Romanian premier universities, the Universitaria Consortium, expressed concerns about Romania's academic system. This is increasingly diverging from international standards (see, for example, M. Miclăuș and O. Micu *Nature* **558**, 189; 2018). In our view, these concerns are being fuelled by government moves that

undermine the status of the country's leading universities, whose resistance to political interference is well known.

At the end of last year, the Ministry of Education hired a commission of foreign experts to reform Romania's university-ranking system. Although insight into the process itself is limited, we were given the commission's first draft report at a public consultation with the universities (see go.nature.com/2lsbwxo for the preliminary version; in Romanian). For example, less than one-quarter of the latest draft's proposed performance indicators seem to accord with those used in major international university rankings; instead, many are used to evaluate institutional facilities such as lecture halls and dormitories. Under this system, all publications would carry equal weight, irrespective of whether they had been peer reviewed.

The draft's criteria seem designed to serve political, not scientific, ends. We are concerned that rewarding small, local universities that have no international standing, rather than those with a record of academic excellence, could foster hierarchies of political influence and further isolate Romania's research community. Rather, Romania needs a ranking system that can accelerate its integration into the international academic arena.

Daniel David, Bálint Markó *Babeș-Bolyai University, Cluj-Napoca, Romania.*
daniel.david@ubbcluj.ro
Competing interests declared (see go.nature.com/2vx72c7 for details).

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the author guidelines and section policies at <http://go.nature.com/cmchno>.

MATERIALS SCIENCE

Glowing fabrics communicate

An approach has been developed for incorporating optoelectronic devices into polymer fibres, which can be woven into fabrics. Such materials could have applications in both telecommunication and health monitoring. [SEE LETTER P.214](#)

WALTER MARGULIS

From time immemorial, textiles have covered our skin and protected it from rain, cold weather and sunlight. The introduction of novel materials and automation techniques widened the use of textiles to carpets, backpacks and car seats. On page 214, Rein *et al.*¹ breathe new life into textiles. The authors present an approach for integrating optoelectronic devices — such as light-emitting diodes — that are commonly used in consumer electronics into fabrics. They demonstrate an optical communication link between two pieces of fabric, and show that their technology can be used to monitor a person's heart rate.

To achieve these feats, Rein and colleagues exploited ready-made, high-quality optoelectronic devices in the form of chips. Such chips are typically a few micrometres in size, and need to be in electrical contact with conducting wires. There are two main challenges to the use of these chips for fabric-based optoelectronics. First, both the chips and the wires need to be protected from the environment — for example, from water. Second, electrical contacts cannot be established for each chip individually, because this would be too costly and time-consuming.

Rein *et al.* addressed the first challenge by embedding the chips and wires in optical fibres made of a transparent polymer. These fibres not only allowed light to be emitted and detected, but also shielded the chips and wires from the environment. The authors then wove the fibres into textiles (Fig. 1). They found that the optoelectronic devices maintained their performance even after ten cycles of a commercial machine-wash.

The authors solved the problem of electrical contacting by means of the method they used to embed the chips and wires in the optical fibres. Optical-fibre fabrication starts with a glass or polymer rod called a preform, which is typically about 2.5 cm in diameter and tens of centimetres in length². The preform is heated in a furnace, and the resulting molten, viscous material is drawn into a fine strand of sub-millimetre diameter: a fibre. Assuming that the conditions are correct, the fibre has a cross-section that replicates that of the preform, but on a much smaller scale. Consequently, two

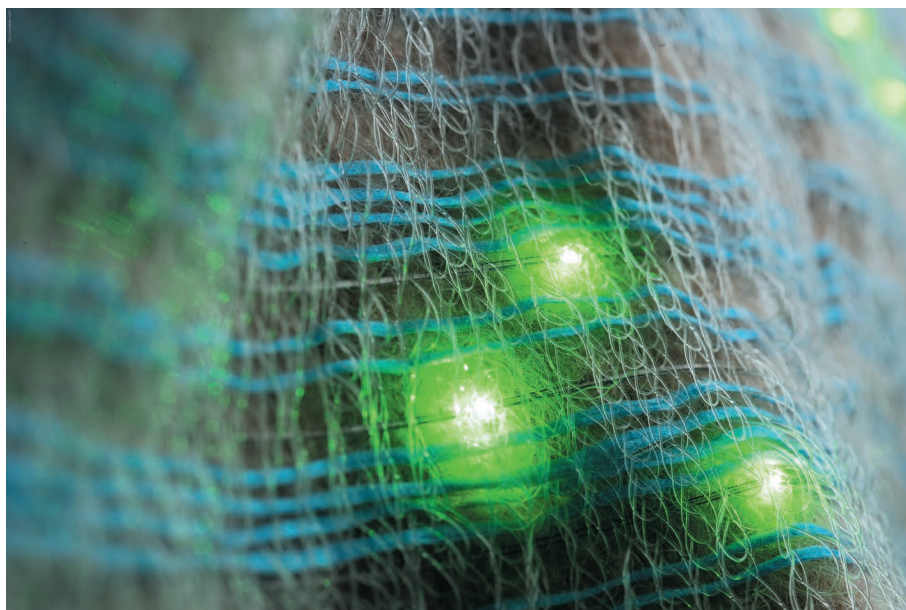


Figure 1 | Optical fibres containing light-emitting diodes integrated into a knitted fabric.

holes separated by 1 mm in the preform could end up as two holes separated by 10 μm in the fibre.

Rein and colleagues discovered that if two fine, hard wires of tungsten or copper are inserted into separate holes in the preform and continuously fed into the preform during fibre drawing³, they can be separated by only a few micrometres in the final fibre. The wires are electrically isolated from each other and are fully encapsulated by the surrounding polymer. Furthermore, a chip that was embedded between the two holes in the preform can end up touching the two wires in the drawn fibre, thus establishing an electrical connection between the chip and the wires. Crucially, individual chips placed near to one another in the preform continue to be operational as a string of devices after fibre drawing. The fabrication technique therefore avoids the need for individual electrical contacting.

Having overcome the challenges of protection and electrical contacting, Rein *et al.* demonstrated potential applications for their fibres. In a beautiful experiment, after mechanically weaving fibres into a textile, the authors lit up many embedded light-emitting diodes of red, green and blue colours. The resulting glowing fabric could be used for

decorative, display or safety purposes. The fabrication technique is also equally suitable for other types of optoelectronic device, such as photodiodes, which generate electrical signals in response to light.

Rein and colleagues used the electrical connections in the fibres both to operate the devices and to convey, through electrical current, information about a device's illumination level. They found that, with one fibre emitting light and one detecting light, an optical communication link could be established. In particular, pulses of light emitted by a fabric at a frequency of up to 3 megahertz could be sensed by a nearby fabric, demonstrating the possibility of transmitting information by optical means.

The authors also explored altering the shapes of the fibres so that they acted as lenses, collimating the light from light-emitting diodes and focusing the light collected by photodiodes. Such alterations improved the efficiency of the demonstrated communication link and increased the maximum distance of the link to about a metre. As a final application, Rein *et al.* show that, if a person presses a finger against a light-emitting fibre and a light-detecting fibre that are near each other, the intensity of the light collected by

GREG HREN

the light-detecting fibre varies according to the person's heart rate⁴. This physiological application of textiles could be used in primary-care settings.

Rein and colleagues' results pave the way for integrating low-cost electronic components into fabrics. Wearable lasers and light detectors and the ability to communicate through garments are some of the possibilities opened up by this work. A strength of the study is the use of high-performance devices that are already available, as opposed to previously reported competing materials and components based on compounds known as chalcogenides⁵ that are still far from reaching the market.

This work describes only the initial phases of

the technique, and much optimization remains to be done. One key step in the fabrication procedure is the mounting of the chips in the preform, which at present is done manually. A mechanized approach could take the technology to a higher level of reproducibility and maturity.

As is the case in many fields, whether or not the technology will enter the market will probably be dictated by economic rather than scientific factors. Nevertheless, practical applications of the fabrics can already be envisaged. Although a high-quality communication link will probably find fierce competition from available technologies, one might more readily expect to see the fabrics

used for a hospital bed sheet to monitor a patient's physiological state, or for a glowing flag in a football stadium. ■

Walter Margulis is in the Department of Fiber Optics, RISE Acreo, 164 40 Kista, Sweden, and in the Department of Applied Physics, Royal Institute of Technology, Stockholm.
e-mail: walter.margulis@ri.se

1. Rein, M. *et al.* *Nature* **560**, 214–218 (2018).
2. Méndez, A. & Morse, T. F. (eds) *Specialty Optical Fibers Handbook* 88 (Academic, 2006).
3. Lee, K., Henry, P., Fleming, S. & Blows, J. L. *IEEE Photon. Technol. Lett.* **18**, 914–916 (2006).
4. Allen, J. *Physiol. Meas.* **28**, R1–R39 (2007).
5. Abouraddy, A. F. *et al.* *Nature Mater.* **6**, 336–347 (2007).

PARASITOLOGY

Drug candidate and target for leishmaniasis

Better treatments are needed for the neglected tropical disease leishmaniasis. The development of a compound that tackles the disease in mice, and the identification of the protein it targets, offer a way forward. SEE ARTICLE P.192

CAROLINA M. C. CATTAPRETA
& JEREMY C. MOTTRAM

The parasite-mediated disease visceral leishmaniasis is prevalent in the tropics and causes 20,000–40,000 deaths across the globe each year¹. The drug treatments currently used for this condition have substantial side effects, are difficult to administer and can result in the evolution of treatment-resistant parasite strains. On page 192, Wyllie *et al.*² present studies of a series of related drug-candidate molecules that are being developed for leishmaniasis treatment. They also identify the target of the most promising compound.

It is more than 100 years³ since drugs based on the chemical element antimony⁴ were first used to treat visceral leishmaniasis, also known as black fever or kala-azar. This sandfly-transmitted disease is caused by the protozoan parasites *Leishmania donovani* in the Old World and *Leishmania infantum* in both the Old World and the New World. Antimony-based drugs are still used today as part of a small range of treatments for the condition. In 2012, the World Health Organization reviewed⁵ the global impact of neglected tropical diseases in the developing world and identified the control of leishmaniasis worldwide and its elimination on the Indian subcontinent as priority targets. To achieve these goals, methods are needed to identify compounds with potent anti-parasitic activity that are suitable for safe and effective therapies.

Parasites from the *Leishmania* genus live

and replicate inside a membrane-bound vacuole in macrophage cells of the immune system. Wyllie and colleagues studied compounds called pyrazolopyrimidines, which are effective against a related protozoan parasite, *Trypanosoma brucei*. The authors optimized the compounds by assessing their

effect on *in vitro* infection of macrophages by *Leishmania* and by testing them using a mouse model of visceral leishmaniasis. They selected one named compound 7 as the best candidate for additional study because it had a good safety profile, high potency and suitable properties for development as an orally administered drug. However, the compound's mode of action was unknown, so the authors sought to identify its molecular target in the parasite. Such target identification is important because it can aid the assessment of possible off-target effects in humans, as well as the likelihood of the emergence of drug resistance.

The authors used a biochemical approach to find proteins that bind to compound 7, and identified three enzymes of interest: CRK3, CRK6 and CRK12. These are similar to cyclin-dependent kinases (CDKs), protein kinases that need to bind a cyclin regulatory protein to enable their kinase activity⁶. When the authors

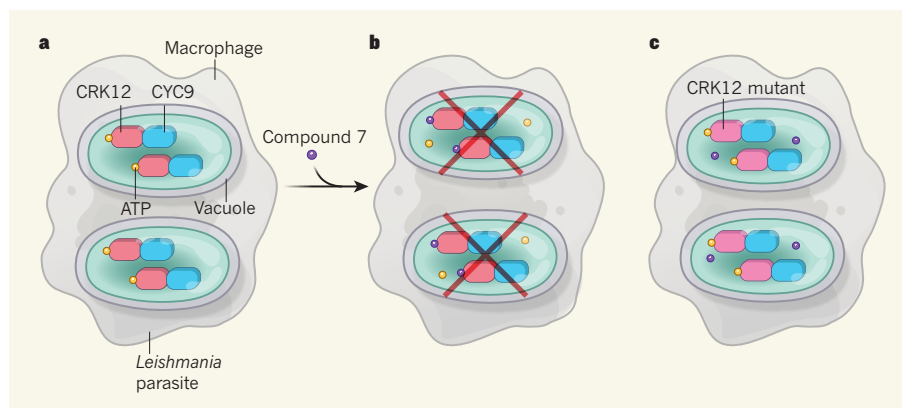


Figure 1 | How a drug candidate targets the *Leishmania* parasite. Wyllie *et al.*² have identified candidate drug molecules for treating the tropical disease leishmaniasis, which they tested in mouse models of the disease. The most promising molecule is called compound 7. **a**, The protozoan *Leishmania* parasite causes leishmaniasis. It infects host immune cells called macrophages and resides in a membrane-bound vacuole. The authors identified the protein kinase enzyme CRK12 as a target of compound 7. This enzyme is similar to cyclin-dependent kinases, has a binding pocket for the molecule ATP and is found in complex with the cyclin protein CYC9. **b**, The level of parasites in mice treated with compound 7 was reduced. The authors' computational modelling studies indicate that compound 7 binds in the ATP-binding pocket of CRK12, thus preventing ATP binding and inhibiting the enzyme's activity, leading to parasite death. **c**, The authors identified a mutation in the catalytic domain of CRK12 that was associated with drug resistance that arose in a laboratory setting. When *Leishmania* parasites were engineered to express this mutant version of CRK12, the effectiveness of compound 7 was reduced. It seems reasonable to speculate that the mutation alters the binding affinity of compound 7 to CRK12, but not that of ATP.

isolated the enzymes, they were found to be associated with cyclin partners. CDKs regulate cell-cycle progression and are key anticancer drug targets⁶.

Wyllie and colleagues investigated parasites that had developed resistance to pyrazolopyrimidine compounds that was caused by exposing parasites to these molecules in the laboratory. The authors carried out whole-genome DNA sequencing to determine how this drug resistance had arisen. These results focused attention on CRK12, together with its associated cyclin protein CYC9, as the probable target of compound 7. This focus was supported by their finding that parasites that express both CRK12 and CYC9 at higher than usual levels have increased resistance to the effects of compound 7. Moreover, the authors identified a mutation in the CRK12 gene in drug-resistant parasites; when this mutation was introduced into wild-type parasites, they became resistant to compound 7 (Fig. 1). The authors' computational modelling studies indicate that compound 7 binds CRK12 in the pocket where the molecule ATP usually binds.

Although CRK12 in complex with CYC9 seems to be the primary molecular target of compound 7, it is possible that other kinases in *Leishmania* could be inhibited by the molecule and contribute to its anti-parasitic activity. The range of protein kinases in *Leishmania* is different from that in humans, so this study provides an impetus to search for other 'druggable' protein kinases in the parasite⁷. Methods such as gene editing⁸ using the CRISPR–Cas9 technique have improved researchers' ability to perform large-scale genetic validation of drug targets in *Leishmania*. However, a major bottleneck in the drug-discovery process for neglected tropical diseases is the identification of highly specific chemical probes that allow chemical validation — evidence that confirms the molecular target of a compound of interest.

The genetic and chemical validation of CRK12 as the target of pyrazolopyrimidines is a key advance because it opens further avenues of exploration for drug discovery. If the pyrazolopyrimidines ultimately fail to be suitable for clinical use, other compounds that inhibit CRK12 could be developed. There have been only a few drug-discovery efforts targeting enzymes in *Leishmania*⁹, mainly because not many targets have been genetically or chemically validated. In this instance, a target-based approach would require the production of CRK12 in complex with CYC9, and the development of an enzyme assay that would be suitable for high-throughput screening to test libraries of chemical compounds. Protein kinases are generally amenable to such approaches, but Wyllie and colleagues report that this has proved challenging so far for CRK12.

Further research should be carried out to investigate the regulation and function of the CRK12–CYC9 complex in *Leishmania* to determine whether modifications such as phosphorylation regulate the activity of

the complex. One key question is why is this complex essential for the survival of *Leishmania* in its mammalian host? The authors found that compound 7 disrupts the parasite's normal cell cycle, which is consistent with the known function of CDKs in cell division. However, the details of the molecular mechanisms at work here remain to be elucidated.

A study¹⁰ in 2016 identified the triazolopyrimidine molecule GNF6702 as having potent activity against *Leishmania*. It acts by inhibiting the cell's proteasomal protein-degradation machinery. Thus, in the past few years, two promising compounds with known targets have emerged. Furthermore, collaborations between pharmaceutical companies, academic institutions and the non-profit Drugs for Neglected Diseases Initiative have identified an increasing number of candidate molecules for leishmaniasis treatment that could be orally administered; these might progress from preclinical studies to clinical trials (see go.nature.com/2lc3mgn).

Is it time to consider testing such chemicals in combination with each other? Combination therapy for visceral leishmaniasis is being evaluated for current drugs because this approach increases treatment efficacy, reduces treatment duration and limits or delays the emergence of drug resistance¹¹. The use of lower concentrations of the compounds and shorter treatment times might help to avoid the emergence

of difficult-to-treat *Leishmania* strains, such as those that have arisen after treatment with the drug miltefosine¹². Wyllie and colleagues' work might open the door for a new drug to be developed. Yet the attrition rate for drug candidates is high. More drug candidates therefore need to be identified to increase the chance that treatments for visceral leishmaniasis will make it to the clinic. ■

Carolina M. C. Catta-Preta and Jeremy C. Mottram are in the Department of Biology, University of York, York YO10 5DD, UK. e-mail: jeremy.mottram@york.ac.uk

1. Alvar, J. et al. *PLoS ONE* **7**, e35671 (2012).
2. Wyllie, S. et al. *Nature* **560**, 192–197 (2018).
3. Vianna, G. *Anais do 7º Congresso Brasileiro de Medicina e Cirurgia* **4**, 426–428 (1912).
4. Brahmachari, U. A *Treatise on Kala Azar* (John Bale, Sons & Danielsson, 1928).
5. World Health Organization. *Research priorities for Chagas disease, human African trypanosomiasis and leishmaniasis* (WHO, 2012).
6. Ferguson, F. M. & Gray, N. S. *Nature Rev. Drug Discov.* **17**, 353–357 (2018).
7. Field, M. C. et al. *Nature Rev. Microbiol.* **15**, 217–231 (2017).
8. Beneke, T. et al. *R. Soc. Open Sci.* **4**, 170095 (2017).
9. Jones, N. G., Catta-Preta, C. M. C., Lima, A. P. C. A. & Mottram, J. C. *ACS Infect. Dis.* **4**, 467–477 (2018).
10. Khare, S. et al. *Nature* **537**, 229–233 (2016).
11. van Griensven, J. et al. *Lancet Infect. Dis.* **10**, 184–194 (2010).
12. Rai, K. et al. *mBio* **4**, e00611–13 (2013).

This article was published online on 25 July 2018.

ALZHEIMER'S DISEASE

Lymphatic waste disposal in the brain

The discovery that a set of lymphatic vessels interacts with blood vessels to remove toxic waste products from the brain has implications for cognition, ageing and disorders such as Alzheimer's disease. [SEE ARTICLE P.185](#)

**MELANIE D. SWEENEY
& BERISLAV V. ZLOKOVIC**

A network of lymphatic vessels acts in tandem with the blood vasculature to regulate fluid balance in the body¹. The brain does not have its own lymphatic network, but the cellular membranes around the brain, known as the meninges, do have a network of lymphatic vessels. This meningeal lymphatic system was first found² in 1787 and has been 'rediscovered' this decade^{3–5}. Do the meningeal lymphatics have a role in brain diseases, as systemic lymphatic vessels do in systemic diseases such as cancer¹? On page 185, Da Mesquita et al.⁶ show that meningeal lymphatic vessels help to maintain both cognitive function and the proper levels of proteins in brain fluids (a process called proteostasis). The finding has

implications for normal ageing and disorders such as Alzheimer's disease.

In the body, lymphatic vessels drain tissues of interstitial fluid (ISF), which contains waste products such as cellular debris and toxic molecules. The ISF forms a protein-rich fluid called lymph that circulates through the lymphatic system back to the circulating blood¹. On its way, lymph is filtered through the lymph nodes, which can initiate immune responses if foreign particles are detected.

The brain does not have its own lymphatic vessels. As such, proteins and waste from the main body of the brain (the parenchyma) are transported within ISF along the walls of blood vessels to reach the cerebrospinal fluid (CSF), which circulates through the meninges⁷. It is well established that proteins, metabolic waste products and other molecules in these fluids can

be removed from the brain by being transported across the walls of blood vessels, thus crossing the blood–brain barrier^{7,8} — a process called transvascular clearance. But it was unknown whether the meningeal lymphatic vessels are also involved in waste clearance.

Da Mesquita *et al.* destroyed the meningeal lymphatic vessels of mice by injecting a vessel-damaging drug into the cisterna magna — a large, CSF-filled space in the meninges. They then administered a fluorescent tracer molecule into the cisterna magna. In mice lacking meningeal lymphatic vessels, the tracer did not reach the deep cervical lymph nodes, to which the meningeal lymphatics normally drain. Similarly, injection of tracers into the brain parenchyma showed reduced ISF drainage into deep cervical lymph nodes. Previous work has shown⁹ that injecting high concentrations of tracer into CSF can cause the diffusion of tracer into the brain along blood vessels — but this transport was also reduced. The authors confirmed these results through several alternative approaches: using different tracers; surgically closing off drainage to the deep cervical lymph nodes; and examining mice genetically engineered so that their lymphatic-vessel development was impaired.

Destruction of the meningeal lymphatics also led to deficits in spatial orientation and fear memory. The brain's hippocampus has a key role in these behaviours, and the researchers found changes in gene expression in this region resembling those seen in neurodegenerative disorders. Collectively, these experiments suggest that drainage of brain ISF and CSF by the meningeal lymphatics is necessary for proper cognitive function.

These findings also raise an interesting question: where did the injected tracers go? One study¹⁰ has shown that tracers injected into the cisterna magna are primarily transported into the blood, and only secondarily into the lymphatic system. Simultaneous measurements of tracer movements into the meningeal lymphatics, other lymphatic vessels (for instance in the neck) and the blood might reveal whether impairment of the meningeal lymphatics leads to a shift in the pathways used to control brain proteostasis, increasing transvascular removal of waste products across the blood–brain barrier (Fig. 1), or their drainage into the venous system in the meninges⁷.

Da Mesquita *et al.* next observed an ageing-induced decrease in the diameter and coverage of meningeal lymphatic vessels, and decreased drainage of tracers from the ISF and CSF into deep cervical lymph nodes. Lymphatic-vessel growth in mice is promoted by a signalling pathway involving vascular endothelial growth factor C (VEGF-C) and its receptor VEGFR3, whereas impairments in the pathway lead to a loss of meningeal lymphatic vessels^{1,3}. Furthermore, treatment with VEGF-C increases the diameter of meningeal lymphatic vessels, improving lymphatic drainage⁴. Consistent with

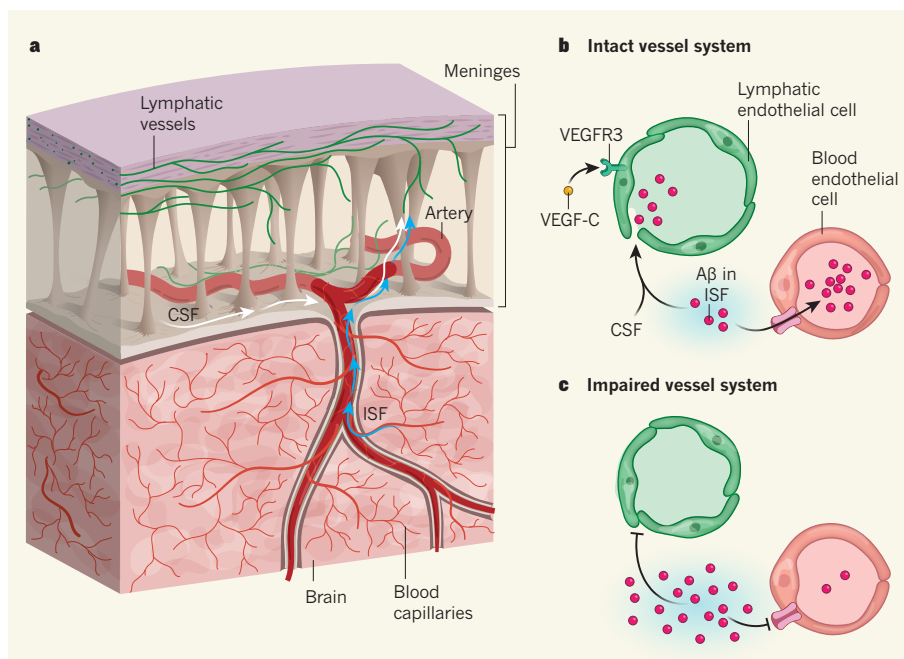


Figure 1 | Regulation of waste clearance in the brain. **a**, The brain does not have its own lymphatic vessels to manage the clearance of waste. Proteins and waste are transported from the brain's interstitial fluid (ISF) along blood-vessel walls to reach the cerebrospinal fluid (CSF) in a space within the meninges — membranes that cover the brain. Da Mesquita *et al.*⁶ report that lymphatic vessels in the meninges drain CSF and ISF containing waste products. **b**, In a healthy mouse brain, lymphatic drainage of both fluids requires signalling between vascular endothelial growth factor C (VEGF-C) and its receptor VEGFR3 on lymphatic endothelial cells lining the wall of meningeal lymphatic vessels. The protein amyloid- β (A β), which is associated with Alzheimer's disease, is primarily removed from the ISF by blood vessels. **c**, During ageing, both vessel systems can become impaired. The diameter of the meningeal lymphatic vessels decreases, causing decreased waste clearance by this route. This defect, along with impaired clearance by blood vessels, leads to A β accumulation in the brain.

these findings, the authors showed that local delivery of the *Vegf-c* gene into the cisterna magna of old mice using a virus restored the drainage of CSF tracer into deep cervical lymph nodes. This change was accompanied by restoration of spatial orientation in old mice.

Age-related impairments in transvascular clearance of waste have been implicated in the accumulation of amyloid- β protein in the brain^{7,11,12} — a hallmark of Alzheimer's disease. Da Mesquita and colleagues investigated the effects of ablating the meningeal lymphatics in two mouse models of Alzheimer's disease, in which amyloid- β protein is produced in neurons and secreted into the ISF. Ablation led to amyloid- β accumulation in the meninges, accelerated amyloid- β deposition in the brain parenchyma and cognitive deficits. The authors also showed that amyloid- β had accumulated in the meninges of people who had Alzheimer's disease, pointing to the potential relevance of these findings for humans.

Notably, the researchers found that the mouse models did not develop any apparent structural or functional changes in the meningeal lymphatics at the time when amyloid- β deposition in the brain parenchyma first became apparent. Viral delivery of *Vegf-c* at this time point could not prevent the cognitive impairments in either model,

suggesting that the early amyloid- β deposition and cognitive impairments in these animals were caused by disruption in another clearance pathway — most likely transvascular clearance. As transvascular-clearance routes gradually deteriorate with age, an increasing burden is probably put on the meningeal lymphatic system. If the capacity of the lymphatic system is reached, this might lead to faulty lymphatic drainage of amyloid- β and other proteins from the ISF and CSF (Fig. 1). Thus, a dynamic relationship between the meningeal lymphatics and blood vessels seems to regulate proteostasis in the brain.

Future work should aim to improve our understanding of waste-clearance pathways from the brain, how the ISF and CSF drain into the meningeal lymphatics, and how these lymphatic vessels interact with the blood vessels at the blood–brain barrier. Such analyses will open up fresh directions for research into cognition, neurodegeneration and Alzheimer's disease. Da Mesquita *et al.* showed that strategies that promote local growth of lymphatic vessels have the potential to improve clearance by meningeal lymphatics to rebuild brain proteostasis, and might lessen amyloid- β deposition. It remains to be determined whether treatments directed at the meningeal lymphatics can also improve the impaired function of blood vessels with age, and whether

enhancing clearance at the blood–brain barrier can improve lymphatic drainage function. ■

Melanie D. Sweeney and Berislav V. Zlokovic are in the Department of Physiology and Neuroscience and at the Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, California 90089–2821, USA.

e-mail: zlokovic@usc.edu

1. Alitalo, K., Tammela, T. & Petrova, T. V. *Nature* **438**, 946–953 (2005).
2. Mascagni, P. *Vasorum lymphaticorum corporis humani historia et ichnographia* (Pazzini Carli, 1787).
3. Aspelund, A. et al. *J. Exp. Med.* **212**, 991–999 (2015).
4. Louveau, A. et al. *Nature* **523**, 337–341 (2015).
5. Absinta, M. et al. *eLife* **6**, e29738 (2017).
6. Da Mesquita, S. et al. *Nature* **560**, 185–191 (2018).

7. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. *Nature Rev. Neurol.* **14**, 133–150 (2018).
8. Zhao, Z. et al. *Nature Neurosci.* **18**, 978–987 (2015).
9. Xie, L. et al. *Science* **342**, 373–377 (2013).
10. Courtice, F. C. & Simmonds, W. J. *Aust. J. Exp. Biol. Med. Sci.* **29**, 255–263 (1951).
11. Shibata, M. et al. *J. Clin. Invest.* **106**, 1489–1499 (2000).
12. Deane, R. et al. *Neuron* **43**, 333–344 (2004).

This article was published online on 25 July 2018.

CONDENSED-MATTER PHYSICS

Stable and switchable polarization in 2D

The electric polarization of materials called ferroelectrics is often suppressed by an internal electric field, limiting uses for these materials. The discovery of a thin-film ferroelectric that is resistant to this field represents a major advance.

TURAN BIROL

Materials known as ferroelectrics have a macroscopic, switchable electric polarization that can be controlled by an external electric field¹. This strong coupling to electric fields, however, is also the bane of ferroelectrics. Electric charges that accumulate on the surfaces of these materials produce an internal electric field called a depolarization field that, if not mitigated by external electrodes, is often large enough to suppress the polarization completely. Writing in *Physical Review Letters*, Xiao et al.² report the observation of ferroelectricity that is invulnerable to the depolarization field in thin films of indium selenide (In_2Se_3). This feature results

from an atypical mechanism that drives the ferroelectricity in indium selenide, and opens the way for both the discovery of other ferroelectrics and further applications for them.

Ferroelectric polarization originates from an asymmetric distribution of atoms in a material's crystal structure — positively charged ions and negatively charged ions are slightly shifted from a symmetric distribution, in opposite directions³ (Fig. 1). However, this arrangement of atoms produces charges on the material's surface, and these charges generate a depolarization field that opposes the polarization. In thin-film ferroelectrics, if the polarization is perpendicular to the plane of the film — the preferable direction for applications — the depolarization field is

usually strong enough to suppress the polarization completely. This suppression limits the potential uses of ferroelectrics in, for example, computer memories⁴ and semiconductor-based electronic devices⁵.

The most commonly studied ferroelectrics are perovskite oxides such as barium titanate (BaTiO_3). In this archetypal ferroelectric, the driving force behind the polarization is the long-range electrostatic (Coulomb) interaction between atoms. Covalent bonds, which involve the sharing of electron pairs between atoms, play a smaller part than the Coulomb interaction in determining the material's ferroelectricity⁶.

Xiao and colleagues instead studied indium selenide, which is a chalcogenide — a compound based on one of the elements in the same group of the periodic table as oxygen. Going down this group, from oxygen to sulfur to selenium, an atom's tendency to attract electrons in a chemical bond towards itself decreases. As a result, bonds have a more strongly covalent character in sulfides and selenides than in oxides, and have a larger effect on the compound's properties.

Indium selenide is a two-dimensional material that consists of five alternating indium and selenium layers, in which the indium–selenium bonds are strongly covalent. Previous theoretical work showed that there are many long-lived atomic configurations of indium selenide that differ in the local bonding environment of the ions in the material's central layers⁷. This work also predicted that the ferroelectric polarization in indium selenide is driven by local covalent bonds, rather than by long-range interactions, and that these bonds are strong enough to prevent the depolarization field from suppressing the polarization — even in thin films that are 3 nanometres thick (equivalent to about three sheets of indium selenide), like those of Xiao and colleagues.

Xiao et al. synthesized their films using both exfoliation (the removal of sheets from a bulk material) and a technique known as van der Waals epitaxial growth, which is an ideal method for growing materials that, like indium selenide, have weakly bound layers⁸. Using imaging tools such as piezoresponse force microscopy, the authors observed a polarization perpendicular to the plane of the film that is stable at temperatures of up to 700 kelvin. They also detected switching of this polarization at room temperature when an electric field was applied.

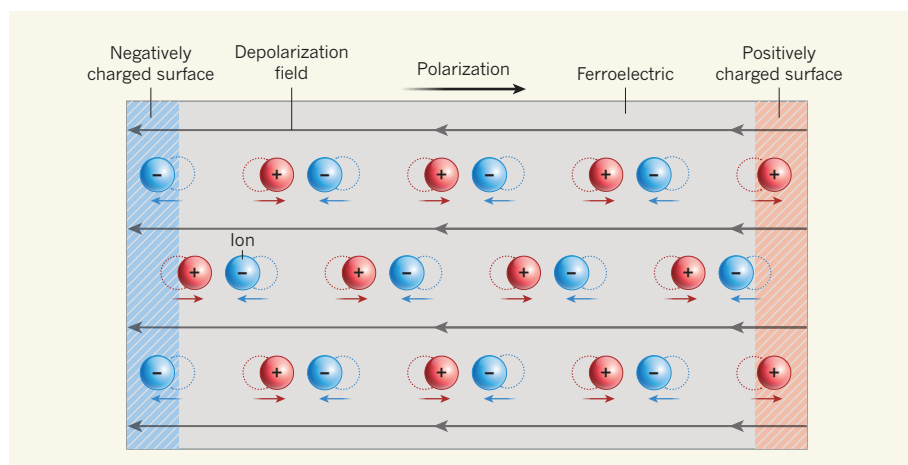


Figure 1 | Ferroelectric polarization. The electric polarization of materials called ferroelectrics originates from the fact that positively charged ions and negatively charged ions are slightly shifted from a symmetric distribution (dotted circles) in opposite directions (coloured arrows). The surfaces of ferroelectrics are negatively or positively charged owing to the presence of unpaired ions. Such charges produce an electric field known as a depolarization field that usually suppresses the polarization, limiting applications for these materials. Xiao et al.² report a ferroelectric in which covalent bonds (not shown) between ions are sufficiently strong that the depolarization field cannot suppress the polarization.

This is not the first report of ferroelectricity in a thin film of a chalcogenide. It is, however, the first observation of an out-of-plane polarization in an atomically thin chalcogenide film that is stable without electrodes mitigating the depolarization field. Such a feature, along with the stability of the polarization at high temperatures, makes indium selenide promising for applications. Now that a chalcogenide has been discovered that has persistent out-of-plane polarization, and in which the mechanism of ferroelectricity is known, we will definitely hear more about chalcogenide ferroelectrics in the coming years.

One previously known group of ferroelectrics that are impervious to the depolarization field are the 'improper' ferroelectrics. In these materials, the emergence of the polarization can be considered to be a side effect of some other structural transition¹. However, rather than being an improper ferroelectric, indium selenide is more likely to be a member of a special group of proper ferroelectrics: the hyperferroelectrics. Such materials have been studied in detail using theoretical approaches⁹, but their polarization has not yet been experimentally shown to be switchable.

Hyperferroelectricity was originally predicted to exist in a group of compounds containing three different elements that, like indium selenide, have a polarization driven by covalent bonds⁹. In these compounds, the Born effective charges (the changes in polarization with respect to the amount by which atoms are displaced) are smaller than those in typical oxide ferroelectrics. As a result, hyperferroelectrics are more resistant to the depolarization field than are their oxide counterparts. So far, indium selenide has not been confirmed as a hyperferroelectric. But if indium selenide were found to be the first hyperferroelectric that contains only two elements, this could lead to the discovery of other 2D chalcogenide ferroelectrics.

Xiao and colleagues' study shows that 2D chalcogenides must be taken seriously in the search for ferroelectrics for technological applications. But it also emphasizes how little is known about the ferroelectricity in this family of materials, compared with the perovskite oxides. The authors' results should also be considered in the context of the increasing interest in the electronic properties of 2D chalcogenides, which can involve exotic phenomena such as quantum spin Hall physics and Weyl semimetals. Future work will surely study the coupling between these phenomena and the polarization, because it could enable different electronic phases to be controlled using electric fields. ■

Turan Birol is in the Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455, USA.

e-mail: tbirol@umn.edu

- of *Ferroelectrics and Related Materials* (Oxford Univ. Press, 2001).
2. Xiao, J. *et al.* *Phys. Rev. Lett.* **120**, 227601 (2018).
3. Rabe, K., Ahn, C. H. & Triscone, J.-M. (eds) *Physics of Ferroelectrics: A Modern Perspective* 1–30 (Springer, 2007).
4. Takashima, D. & Kunishima, I. *IEEE J. Solid State Circuits* **33**, 787–792 (1998).
5. Hoffman, J. *et al.* *Adv. Mater.* **22**, 2957–2961 (2010).

6. Cohen, R. E. & Krakauer, H. *Phys. Rev. B* **42**, 6416–6423 (1990).
7. Ding, W. *et al.* *Nature Commun.* **8**, 14956 (2017).
8. Walsh, L. A. & Hinkle, C. L. *Appl. Mater. Today* **9**, 504–515 (2017).
9. Garrity, K. F., Rabe, K. M. & Vanderbilt, D. *Phys. Rev. Lett.* **112**, 127601 (2014).

This article was published online on 30 July 2018.

MATERIALS SCIENCE

Designer topology in graphene nanoribbons

In materials known as graphene nanoribbons, topological states can be precisely engineered and probed, providing an experimental platform for studying electronic topology. [SEE LETTERS P.204 & P.209](#)

KATHARINA J. FRANKE & FELIX VON OPPEN

For more than a decade, two-dimensional sheets of carbon atoms known as graphene have captured researchers' imaginations. Last year, it was predicted that electronic states in narrow strips of graphene — dubbed graphene nanoribbons — could have different topologies depending on the width of the strip¹. On pages 204 and 209, respectively, Rizzo *et al.*² and Gröning *et al.*³ report experiments that confirm this prediction. Their results show that graphene nanoribbons provide a flexible and highly precise platform for designing and fabricating materials that have what is known as a non-trivial topology. The authors suggest that such materials could be

used to realize desired exotic topological states for quantum technologies.

We learn in school that materials can differ starkly in their electrical properties. The difference between conductors and insulators is rooted in the states that are available to the electrons in these materials. In conductors, such as metals, electrons can move freely because available states exist at arbitrarily low energies. By contrast, the electrons in insulators are effectively localized, and do not conduct electricity unless they are provided with sufficient energy to overcome an energy gap.

This understanding of conductors and insulators was an early triumph for the application of quantum theory to materials. However,

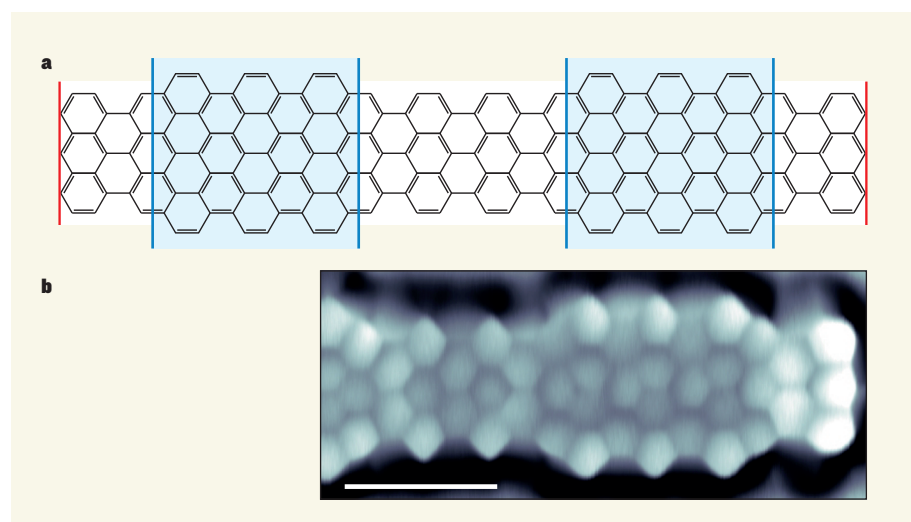


Figure 1 | A graphene nanoribbon. **a**, Rizzo *et al.*² and Gröning *et al.*³ synthesized strips of graphene (a two-dimensional form of carbon) known as graphene nanoribbons (black). The nanoribbons alternated in width such that the topologies of electronic states in the narrow sections (white) and wide sections (blue) were trivial and non-trivial, respectively. The authors report localized topological electronic states at the junctions (blue lines) between narrow and wide sections, and at the ends (red lines) of the nanoribbons. **b**, This micrograph shows one end of a nanoribbon studied by Rizzo and colleagues. Scale bar, 1 nanometre.

1. Lines, M. E. & Glass, A. M. *Principles and Applications*

over the past decade or so, researchers have learnt that this picture needs to be amended in fundamental ways. This realization has led to the discovery of materials known as topological insulators, which are insulating in their interior but robustly conducting on their boundaries⁴. Correspondingly, these materials have an energy gap in their interior, but are gapless on their boundaries. This behaviour reflects beautiful, albeit somewhat abstract, topological properties of the materials' electronic states.

Rizzo *et al.* and Gröning *et al.* have experimentally demonstrated that graphene nanoribbons can be used to produce such topological states. Defect-free graphene nanoribbons can be grown on metallic substrates in a remarkably flexible manner⁵. Starting with cleverly designed precursor molecules, the nanoribbons' terminations and widths can be controlled with single-atom precision. The authors used this synthesis technique to grow nanoribbons that alternate in width (Fig. 1a).

The widths were chosen such that the nanoribbons consist of alternating topologically trivial and non-trivial segments. Whenever two materials of different topology are brought into contact, gapless states must form at the interface. Consequently, such states are produced at the junctions between the nanoribbon segments. Because nanoribbons are essentially one-dimensional, each of these gapless junction states is simply an individual electron orbital localized in the vicinity of the intersection.

But the topology of the nanoribbons does not stop here. Rizzo *et al.* and Gröning *et al.* used the junction states as building blocks to engineer yet another system. This system is closely related to an archetypal model of electronic topology known as the Su–Schrieffer–Heeger (SSH) model, which emerged in the late 1970s from the study of organic conductors such as polyacetylene⁶.

Although the SSH model is simple, it has remarkable properties. In particular, a finite chain of electronic orbitals described by the SSH model can have gapless topological states localized at its ends. The crucial ingredient in the model is an alternation of weak and strong bonds between neighbouring electron orbitals.

In the authors' nanoribbons, adjacent gapless junction states straddle narrow or wide regions of the material. The coupling of these states is stronger across the wide regions than across the narrow regions, producing exactly the bond alternation that underlies the SSH model. Such a coupling is therefore expected to generate topological states at the ends of the nanoribbons, assuming that these materials are suitably designed¹.

Rizzo *et al.* and Gröning *et al.* confirmed this theoretical prediction to an impressive degree. The authors used a combination of scanning tunnelling microscopy and spectroscopy to probe and visualize the electronic properties of the nanoribbons with

atomic-scale spatial resolution (Fig. 1b). They observed the junction states — which formed broadened energy bands as a result of their coupling — and the end states associated with the bond alternation. Of note is the fact that the authors grew and probed their nanoribbons on a highly conducting gold substrate, which effectively weakens the electric forces between the electrons in the nanoribbons. Without such a conducting substrate, these forces could be substantial, and might produce additional interesting physics¹.

Beyond fabricating these specific nanoribbons and exploring their electronic topologies, the two studies reveal several key insights. For instance, the production of topological electronic materials is often hampered by sample imperfections. Frequently, defects induce a large internal conductivity, even if the material is nominally a topological insulator. This problem is particularly severe in 1D systems, in which the electrons cannot circumvent defects. Such systems are often fabricated using a top-down approach, in which the materials are patterned from larger structures. A promising avenue for alleviating the issue of sample imperfections is to produce the systems by means of a bottom-up method, such as that

used by the authors, in which the materials are made by chemical processes.

These studies also highlight the potential of using topological boundary states for materials engineering. This idea can be extended to higher dimensions than the authors' 1D system, for instance to periodic 'superlattices' made of alternating topologically trivial and non-trivial layers. Finally, the authors suggest that, when in contact with a superconductor, the nanoribbons could act as a topological superconductor — another fascinating class of topological electronic state that might have applications in quantum computing. ■

Katharina J. Franke and Felix von Oppen
are in the Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany.
e-mails: katharina.franke@physik.fu-berlin.de;
vonoppen@physik.fu-berlin.de

1. Cao, T., Zhao, F. & Louie, S. G. *Phys. Rev. Lett.* **119**, 076401 (2017).
2. Rizzo, D. J. *et al.* *Nature* **560**, 204–208 (2018).
3. Gröning, O. *et al.* *Nature* **560**, 209–213 (2018).
4. Hasan, M. Z. & Kane, C. L. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
5. Cai, J. *et al.* *Nature* **466**, 470–473 (2010).
6. Su, W. P., Schrieffer, J. R. & Heeger, A. J. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).

IMMUNOLOGY

Making mitochondrial DNA is inflammatory

Activation of the inflammasome protein complex in immune cells is a key step that triggers an innate immune response. It emerges that the synthesis and oxidation of mitochondrial DNA drives this activation step. [SEE ARTICLE P198](#)

MICHAEL P. MURPHY

The innate immune response mounts a defence when immune cells recognize general hallmarks of infection, such as lipopolysaccharide (LPS) molecules, which are present in many types of bacterium. However, the inappropriate unleashing of an innate immune response can lead to autoimmune disorders. Gaining a better understanding of how innate immune responses are regulated might lead to improvements in clinical treatments for such disorders. On page 198, Zhong *et al.*¹ report that DNA synthesis in organelles called mitochondria has a key role in triggering an innate immune response.

Mitochondria can regulate how immune cells respond to infection and tissue damage. For example, these organelles can produce pro- or anti-inflammatory signals by altering the levels of metabolites produced in the Krebs cycle^{2,3}, or by changing the level of production of reactive oxygen species (ROS)^{4,5}.

More and more examples are being found of mitochondrial functions being repurposed in unexpected ways to contribute to inflammatory signalling^{2–5}.

The inflammasome is a multiprotein complex that assembles in immune cells during an innate immune response. It provides defensive functions when the inflammasome-associated enzyme caspase-1 cleaves and activates inflammatory proteins such as IL-1 β . Inflammasomes that contain the protein NLRP3 can form in immune cells called macrophages, and the initial steps in the assembly or priming of this type of inflammasome are reasonably well understood: if LPS binds to the receptor protein TLR4 on the macrophage surface, there is an increase in signalling by the NF- κ B pathway. This causes an increase in expression of NLRP3 and of the precursor form of IL-1 β .

However, the process that triggers inflammasome activation, which occurs when the enzyme caspase-1 is recruited to

the inflammasome and aids the production of inflammatory proteins, is not fully understood. It was puzzling that many highly diverse molecular cues can trigger this step. Yet hints from experimental studies have suggested that these cues might ultimately act through a mitochondrial pathway associated with high levels of mitochondrial ROS^{3,6,7} — which are required to oxidize mitochondrial DNA — and the release of oxidized mitochondrial DNA, which binds to the inflammasome⁸.

The binding of mitochondrial DNA to an NLRP3-containing inflammasome is essential for inflammasome activation^{9,10}. Zhong and colleagues studied mice to assess whether the availability of this mitochondrial DNA might regulate inflammation. The authors engineered animals so that their immune cells lack the protein TFAM, which is required for mitochondrial DNA replication. This led to a loss of mitochondrial DNA, resulting in defective inflammasome activation. When the authors transferred synthetic oxidized mitochondrial DNA into macrophage cells grown *in vitro* from the animals lacking TFAM, this triggered inflammasome activation in response to an LPS signal.

The authors investigated how mitochondrial sensing of innate-immunity triggers might lead to mitochondrial-DNA synthesis. They report that LPS binding to TLR4 activates a pathway that drives expression of the enzyme CMPK2, which is required to produce the nucleotide cytidine triphosphate (CTP) (Fig. 1). Zhong and colleagues engineered mouse macrophage cells to lack CMPK2, and found that such cells were deficient in inflammasome activation. It is unknown how CMPK2 and the mitochondrial CTP pool operate as a control point for mitochondrial-DNA synthesis in macrophages.

To track newly made mitochondrial DNA, the authors introduced a labelled building block of DNA into macrophage cells grown *in vitro*. When these cells received an inflammasome-activating cue, such as LPS, newly made DNA was found to be associated with the inflammasome, and DNA-sequence analysis confirmed its mitochondrial origin. Intriguingly, the authors did not find evidence that the oxidized DNA had to be mitochondrial DNA to bind to the inflammasome. The introduction of oxidized nuclear DNA could do the job just as well, suggesting that oxidized DNA is the key signal.

Zhong and colleagues' work fills in the gap between the priming and activation of the inflammasome by indicating that newly synthesized mitochondrial DNA can give rise to oxidized mitochondrial DNA fragments that exit the organelle to activate NLRP3-containing inflammasomes. Their core conclusions are convincing; however, the solidity of these findings inevitably focuses our attention on those points that are still uncertain. One intriguing issue is the nature of the newly synthesized mitochondrial DNA. The authors'

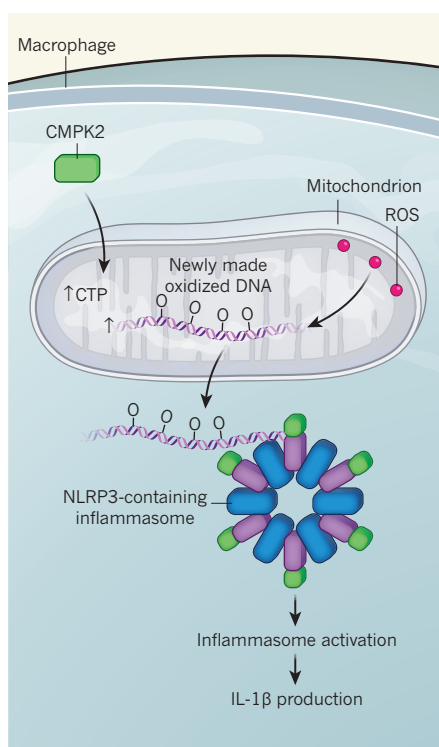


Figure 1 | Newly synthesized oxidized mitochondrial DNA triggers inflammasome activation. The inflammasome is a multiprotein complex that has a key role in generating a defence response and is found in immune cells such as macrophages. How inflammasomes that contain the protein NLRP3 are activated was not fully understood. Zhong *et al.*¹ studied inflammasome activation in mice and report that, when macrophages sense a foreign molecular cue, levels of the enzyme CMPK2 increase. CMPK2 localizes to an organelle called a mitochondrion and drives an increase in the levels of the nucleotide cytidine triphosphate (CTP). This event is linked to synthesis of mitochondrial DNA, and this freshly generated DNA is thought to be oxidized (O denotes oxidized DNA) by reactive oxygen species (ROS). The authors find that oxidized DNA exits the mitochondrion, binds to the NLRP3-containing inflammasome and activates it. This leads to the production of inflammatory proteins such as IL-1 β .

findings suggest that this is produced by the polymerase enzyme that normally replicates mitochondrial DNA, but it is unclear whether the entire mitochondrial DNA sequence is replicated or whether replication terminates prematurely once sufficient DNA is made to generate an inflammatory signal. And is newly formed mitochondrial DNA particularly susceptible to oxidative damage? Could it be that the newly synthesized DNA lacks protection from the nucleoid proteins that normally bind to mitochondrial DNA, thereby increasing its exposure to ROS?

The authors incorporated the oxidized nucleotide 8-hydroxy-2'-deoxyguanosine into cells grown *in vitro* as a way to generate oxidized mitochondrial DNA. This type of nucleotide is frequently found in oxidized DNA, but there are many other types of oxidative DNA modification, and it would

be interesting to explore which of these can activate inflammasomes.

How do the ROS needed for DNA oxidation arise? The tacit assumption is that non-specific organelle damage generates ROS. Yet this is debatable¹¹. I suspect that the mitochondrial ROS production during NLRP3-inflammasome activation might be just as regulated as the process of mitochondrial DNA synthesis. Perhaps the succinate molecules that accumulate after LPS stimulation are oxidized to drive mitochondrial ROS production⁴.

Another area worthy of future investigation is how oxidized mitochondrial DNA is released into the cytoplasm. The authors make the plausible proposal that a large mitochondrial pore might provide an exit route. One candidate is the mitochondrial permeability transition pore, which forms in response to increased levels of ROS¹². However, there are also other possibilities to consider: for example, mitochondria can release microvesicles containing oxidized DNA and protein¹³.

The authors' insights into the activation of NLRP3-containing inflammasomes immediately suggest targets for the development of anti-inflammatory drugs. One area to explore is inhibition of CMPK2 during inflammation, and other parts of the pathway that the authors uncovered are worth considering as targets, too.

This finding of yet another fascinating link between mitochondria and inflammatory signalling in the innate immune system might reflect the organelle's early evolutionary origins as a bacterial cell. This inherent otherness could give mitochondria a head start in being recognized as foreign by the innate immune system.

On page 238, Dhir *et al.*¹⁴ report that the release of double-stranded RNA from mitochondria acts as an antiviral signal. This provides an additional example that the release of mitochondrial nucleic acids to the cytoplasm can act as a signal that triggers a defence response. ■

Michael P. Murphy is at the Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge CB2 0XY, UK.
e-mail: mpm@mrc-mbu.cam.ac.uk

1. Zhong, Z. *et al.* *Nature* **560**, 198–203 (2018).
2. Bambouskova, M. *et al.* *Nature* **556**, 501–504 (2018).
3. Mills, E. L. *et al.* *Nature* **556**, 113–117 (2018).
4. Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. *Nature* **469**, 221–225 (2011).
5. Mills, E. L. *et al.* *Cell* **167**, 457–470.e13 (2016).
6. He, Y., Hara, H. & Núñez, G. *Trends Biochem. Sci.* **41**, 1012–1021 (2016).
7. Schroder, K. & Tschopp, J. *Cell* **140**, 821–832 (2010).
8. Shimada, K. *et al.* *Immunity* **36**, 401–414 (2012).
9. Zhong, Z. *et al.* *Cell* **164**, 896–910 (2016).
10. Nakahira, K. *et al.* *Nature Immunol.* **12**, 222–230 (2011).
11. Murphy, M. P. *et al.* *Cell Metab.* **13**, 361–366 (2011).
12. Bernardi, P. & Di Lisa, F. *J. Mol. Cell. Cardiol.* **78**, 100–106 (2015).
13. Sugiura, A., McLelland, G.-L., Fon, E. A. & McBride, H. M. *EMBO J.* **33**, 2142–2156 (2014).
14. Dhir, A. *et al.* *Nature* **560**, 238–242 (2018).

This article was published online on 25 July 2018.

Coherent spin–photon coupling using a resonant exchange qubit

A. J. Landig^{1,4*}, J. V. Koski^{1,4}, P. Scarlino¹, U. C. Mendes², A. Blais^{2,3}, C. Reichl¹, W. Wegscheider¹, A. Wallraff¹, K. Ensslin¹ & T. Ihn¹

Electron spins hold great promise for quantum computation because of their long coherence times. Long-distance coherent coupling of spins is a crucial step towards quantum information processing with spin qubits. One approach to realizing interactions between distant spin qubits is to use photons as carriers of quantum information. Here we demonstrate strong coupling between single microwave photons in a niobium titanium nitride high-impedance resonator and a three-electron spin qubit (also known as a resonant exchange qubit) in a gallium arsenide device consisting of three quantum dots. We observe the vacuum Rabi mode splitting of the resonance of the resonator, which is a signature of strong coupling; specifically, we observe a coherent coupling strength of about 31 megahertz and a qubit decoherence rate of about 20 megahertz. We can tune the decoherence electrostatically to obtain a minimal decoherence rate of around 10 megahertz for a coupling strength of around 23 megahertz. We directly measure the dependence of the qubit–photon coupling strength on the tunable electric dipole moment of the qubit using the ‘AC Stark’ effect. Our demonstration of strong qubit–photon coupling for a three-electron spin qubit is an important step towards coherent long-distance coupling of spin qubits.

The ability to transmit quantum information over long distances is desirable for quantum information processors¹. Circuit quantum electrodynamics provides a well-established platform for connecting distant qubits²: microwave photons in a superconducting waveguide resonator couple to the electric dipole moment of multiple qubits, which are fabricated close to the resonator. Strong qubit–photon coupling has been realized with superconducting qubits³ and, recently, the coherence properties of charge qubits in semiconductor quantum dots have improved sufficiently to enable strong coupling^{4–6}. Even better coherence is expected by transferring the quantum information from electron charge to spin^{7,8}. However, this approach comes with a major challenge because the coupling of photons to spins is several orders of magnitude weaker than their coupling to charge⁹. This challenge can be overcome by introducing an electric dipole moment to the spin states. For single-electron spin qubits, spin and charge are coupled by using materials with strong spin–orbit coupling¹⁰, devices with ferromagnetic leads¹¹ or a magnetic-field gradient generated by an on-chip micromagnet^{12–14}. A different approach is realized in the resonant exchange qubit^{15–19}, in which the spin exchange interaction couples two states with an equal three-electron charge distribution and equal total spin, but different spin arrangements. This interaction also gives rise to an electrical dipole moment that enables coherent qubit–photon coupling. Here, we implement such a three-electron spin qubit in a circuit quantum electrodynamics architecture^{20,21} hosted in GaAs and achieve strong spin–photon coupling, as evident from the observation of vacuum Rabi mode splitting. Both the spin decoherence and the qubit–photon coupling strength can be controlled electrostatically²².

Quantum device

In Fig. 1a, b we show optical and scanning electron micrographs of our hybrid quantum device. Electrons are trapped in a triple-quantum-dot structure by electrostatic confinement created by gold gates (Fig. 1b) on top of a GaAs/AlGaAs heterostructure. The heterostructure hosts a two-dimensional electron gas 90 nm

below the surface of the triple-quantum-dot region, which has a mobility of $\mu = 3.2 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and an electron density of $n_e = 2.2 \times 10^{11} \text{ cm}^{-2}$ at 4.2 K. The electrostatic potentials of the left, middle and right quantum dots are tuned using the respective plunger-gate voltages V_L , V_M and V_R . A quantum point contact acts as a charge sensor that allows us to determine the charge configuration of the triple quantum dot. We operate the triple quantum dot as a three-electron spin qubit¹⁹, as discussed in detail below.

To couple the qubit to microwave photons, the plunger gate of the left quantum dot extends to the superconducting microwave resonator (Fig. 1a). The left plunger gate is also DC-biased via a resistive gold line, which is connected to the field anti-node of the centre conductor of the resonator. The coupling strength g_s between qubit and resonator photons is proportional to the square root of the characteristic impedance of the resonator ($\sqrt{Z_r}$)^{6,23}. It is enhanced by fabricating the resonator, as shown in Fig. 1a, from a thin (about 15 nm) and narrow (roughly 300 nm) centre conductor made of the high-kinetic-inductance material NbTiN²⁴. We estimate $Z_r = \sqrt{L_1/C_1} \approx 1.3 \text{ k}\Omega$, with $L_1 \approx 150 \mu\text{H m}^{-1}$ ($C_1 \approx 90 \text{ pF m}^{-1}$) the inductance (capacitance) of the resonator per unit length, which results in an enhancement in the coupling strength by a factor of five compared to a standard impedance-matched $Z_r = 50 \Omega$ resonator. Our choice of material and design allows us to operate the resonator in the presence of an external magnetic field applied parallel to the plane of the resonator²⁴. In the experiments described here, we apply a magnetic field of $B_{\text{ext}} = 200 \text{ mT}$.

Strong spin–qubit–photon coupling

To demonstrate strong coupling of the spin qubit with a microwave photon, we first detune the transition frequency of the qubit from the resonance frequency of the resonator. In this detuned situation, we determine a resonator resonance frequency of $\nu_r = 4.38 \text{ GHz}$ and a line width of $\kappa/(2\pi) = 47.1 \text{ MHz}$ at an average photon occupation of less than 1 (see inset of Fig. 1c). When the spin qubit is tuned into resonance with the resonator, we observe two distinct peaks in the

¹Department of Physics, ETH Zürich, Zurich, Switzerland. ²Institut quantique and Département de Physique, Université de Sherbrooke, Sherbrooke, Quebec, Canada. ³Canadian Institute for Advanced Research, Toronto, Ontario, Canada. ⁴These authors contributed equally: A. J. Landig, J. V. Koski. *e-mail: alandig@phys.ethz.ch

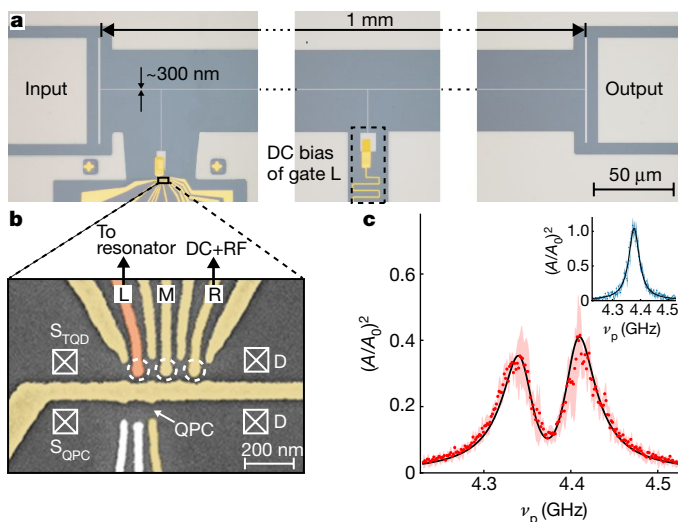


Fig. 1 | Hybrid quantum device and vacuum Rabi splitting. **a**, Optical micrograph of the device split into three parts, showing the resonator, which is capacitively coupled to the input and output transmission lines. The region for the DC bias of gate L (see **b**) that connects to the centre of the resonator is indicated as a dashed black rectangle. **b**, False-colour scanning electron micrograph of the gate structure defined by electron beam lithography. The two white gates are kept at zero voltage in our experiments. The gate highlighted in orange is electrically connected to the resonator. The approximate positions of the left, middle and right quantum dots are indicated by dashed white circles; their corresponding plunger gates are labelled 'L', 'M' and 'R'. The right plunger gate is biased with both DC and microwave (RF) signals. The triple quantum dot and quantum point contact (QPC) have separate ohmic source contacts (S_{TQD} and S_{QPC}) and a common drain contact (D). **c**, Resonator transmission $(A/A_0)^2$ as a function of resonator probe frequency ν_p for the uncoupled (blue, inset) and coupled (red, main plot) configuration, showing vacuum Rabi mode splitting as a result of strong spin–photon coupling. The standard deviation of repeated measurements is indicated by the shaded region. The qubit parameters for the coupled configuration are specified in Fig. 4. The solid black lines are fits to an input–output model²⁵.

transmission spectrum (Fig. 1c). This splitting of the resonance of the resonator into two well-separated peaks, known as vacuum Rabi mode splitting, is the characteristic signature of strong coherent hybridization of a single microwave photon in the resonator and the spin qubit in the triple quantum dot. From a fit of the vacuum Rabi splitting to an input–output model²⁵, we extract a qubit–photon coupling strength of $g_s/(2\pi) = 31.4 \pm 0.3$ MHz and a qubit decoherence rate of $\gamma_2/(2\pi) = 19.6 \pm 0.5$ MHz. These values confirm that our quantum device operates in the strong coupling regime, which is supported by the fact that the approximate peak separation is larger than the widths of peaks, $2g_s > \kappa/2 + \gamma_2$. This is our main result; we provide more details on how it was achieved below.

Triple-quantum-dot spin qubit

The spin qubit is formed by tuning the triple quantum dot into the three-electron regime. In Fig. 2a we show the charge stability diagram of the triple quantum dot, as measured by the charge detector. Regions with different charge configurations (k, l, m) are indicated, where the integers k, l and m express the number of electrons in the three dots. The qubit operation point is located in the narrow (1, 1, 1) region between the (2, 0, 1) and (1, 0, 2) regions. As illustrated in Fig. 2b, we introduce an asymmetry parameter ε and a detuning parameter Δ to quantify differences in the energies $E(i)$ of the three relevant charge configurations i in the absence of interdot tunnelling: $\varepsilon = [E(2, 0, 1) - E(1, 0, 2)]/2$ and $\Delta = E(1, 1, 1) - [E(2, 0, 1) + E(1, 0, 2)]/2$. Both parameters are tuned experimentally using the plunger-gate voltages: ε increases by increasing V_L and decreasing V_R , whereas Δ increases by increasing V_L and

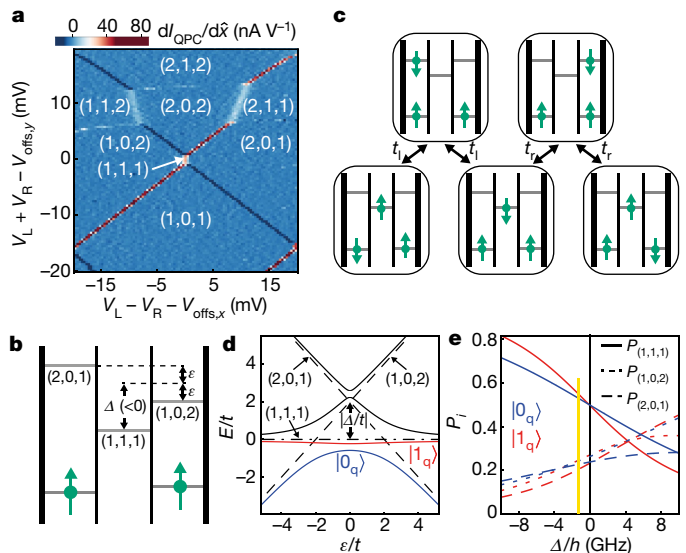


Fig. 2 | Spin-qubit operation regime. **a**, Differential quantum point contact current $dI_{\text{QPC}}/d\hat{x}$, where $\hat{x} = V_L - V_R - V_{\text{offs},x}$, as a function of different combinations of plunger gate voltages V_L and V_R . $V_{\text{offs},x}$ and $V_{\text{offs},y}$ are voltage offsets in the x and y directions. **b**, Schematic of the triple quantum dot, defining the asymmetry and detuning parameters ε and Δ , respectively. The three grey lines indicate the possible energy levels for the addition of the third electron. **c**, Illustration of the three electron states in the triple quantum dot that form the spin qubit. The states mix via tunnel couplings t_l and t_r . **d**, Eigenenergies E/t of the system illustrated in **c** as a function of ε/t for $\Delta/t = -2$ and symmetric tunnel coupling $t_l = t_r = t$. Dashed lines indicate the energy of the charge states (2, 0, 1) and (1, 0, 2) for $t_l = t_r = 0$. The dash-dotted line is the eigenenergy of the $S = 3/2$, $S_z = 1/2$ state, which does not couple to any of the other states (Supplementary Information, section S1). This line also corresponds to the energy of the (1, 1, 1) states for $t_l = t_r = 0$. The spin-qubit states $|0_q\rangle$ (blue) and $|1_q\rangle$ (red) are highlighted. **e**, Probabilities $P_{(1,1,1)}$ (solid lines), $P_{(2,0,1)}$ (dashed lines) and $P_{(1,0,2)}$ (dotted lines), as defined in the main text, for $|0_q\rangle$ (blue) and $|1_q\rangle$ (red) as a function of Δ/h . The plot is obtained for $t_l/h = 9.04$ GHz, $t_r/h = 7.99$ GHz and $\varepsilon/h = -1.03$ GHz. The position in Δ/h at which Fig. 1c was recorded is indicated by the yellow line.

V_R while decreasing V_M . Other charge configurations are not relevant, because the charging energies of the quantum dots are of the order of 1 meV (240 GHz), much larger than the thermal energy $k_B T = 3 \mu\text{eV}$ (620 MHz) for our experiments, which were performed at an electronic temperature of $T = 30$ mK (and where k_B is the Boltzmann constant).

In general, there are eight different spin configurations for three spins. For the asymmetric charge configurations (2, 0, 1) and (1, 0, 2), the three triplet states within the doubly occupied dots do not play a part because the singlet–triplet splitting of roughly 1 meV (240 GHz) is much larger than the temperature⁷. This leaves us with two relevant spin configurations for each of the two asymmetric charge configurations. Two of them, each with a z component of total spin of $S_z = 1/2$, are depicted in the top row of Fig. 2c. The other two are obtained by flipping the spin in the singly occupied dot, giving $S_z = -1/2$. These spin configurations of the asymmetric charge configurations couple by tunnelling to the spin configurations of the (1, 1, 1) charge configuration. The qubit states are formed by a coherent superposition of the five basis states with $S_z = 1/2$ (Fig. 2c)¹⁹. An equivalent set of basis states with $S_z = -1/2$, which differs only in the Zeeman energy, exists but is not depicted. Mixing of these different S_z states by an Overhauser field of about 5 mT⁷ is suppressed by the much larger externally applied magnetic field. The (1, 1, 1) states couple via the exchange interaction between electrons in neighbouring dots: an electron in the middle dot can be exchanged with an electron of opposite spin in the left or right dot by tunnelling to the asymmetric charge state.

We do not consider the (1, 1, 1) state with $S_z = 3/2$ because, for our choice of external magnetic field ($B_{\text{ext}} = 200$ mT), its energy is more

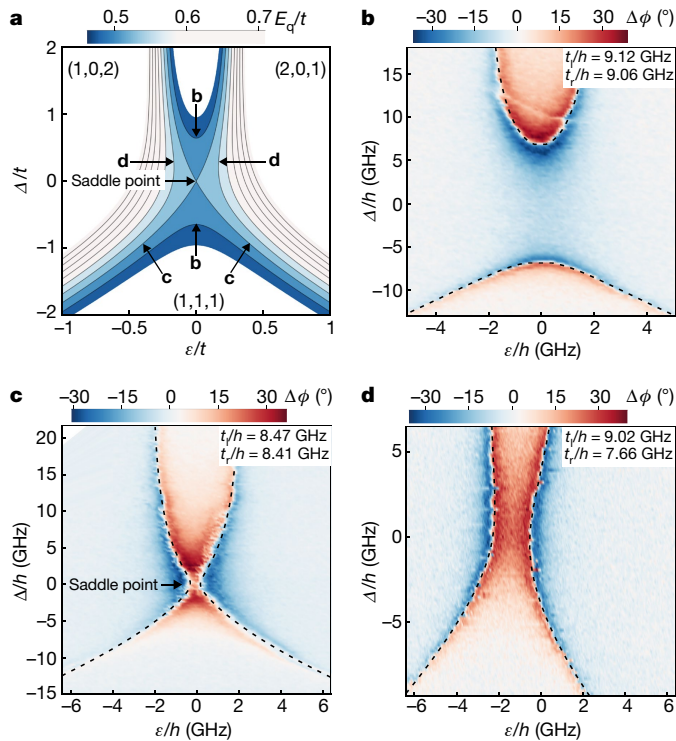


Fig. 3 | Resonator response. **a**, Contour plot of the normalized qubit energy E_q/t for symmetric tunnel coupling t as a function of the asymmetry and detuning parameters ε/t and Δ/t . The energy contours probed in **b–d** are labelled, as is the saddle point in the qubit energy. The energetically favoured three-electron charge configurations are also indicated. **b–d**, Phase difference $\Delta\phi$ of the signal transmitted through the resonator, measured on-resonance for different tunnel-coupling configurations t_l and t_r . The dashed lines indicate fits to the theoretical qubit energy contours (see **a**).

than $h \times 1$ GHz higher than the excited-state energy of the qubit (where h is the Planck constant). It therefore does not form the ground state of the system and does not coherently couple via fluctuations in the Overhauser field to the qubit states. The $S_z = 3/2$ state becomes relevant for $B_{\text{ext}} \geq 1$ T (see Supplementary Information, section S3).

The two lowest-energy eigenstates of the system define the ground $|0_q\rangle$ and the excited $|1_q\rangle$ state of the qubit, which has energy $E_q(\Delta, \varepsilon, t_l, t_r) = E_{|1_q\rangle} - E_{|0_q\rangle}$, where $t_{l(r)}$ is the tunnel coupling between the middle dot and the left (right) dot (see Fig. 2d). In the limit $\Delta \ll -t_{l,r}$, the qubit states predominantly have the same charge configuration— $(1, 1, 1)$ —and are given by $|0_q\rangle \approx |0\rangle \equiv (|\uparrow, \uparrow, \downarrow\rangle - |\downarrow, \uparrow, \uparrow\rangle)/\sqrt{2}$ and $|1_q\rangle \approx |1\rangle \equiv (2|\uparrow, \downarrow, \uparrow\rangle - |\uparrow, \uparrow, \downarrow\rangle - |\downarrow, \uparrow, \uparrow\rangle)/\sqrt{6}$ (ref. 19; Supplementary Information, section S1). Because both qubit states have the same total spin of $1/2$, the finite qubit energy is not determined by an external magnetic field but by the exchange interaction (which is proportional to t^2/Δ) between the $|0\rangle$ and $|1\rangle$ spin states, thus realizing the resonant exchange qubit. In this regime, the qubit is minimally influenced by charge noise, but also couples weakly to photons. In the other extreme ($\Delta \gg t_{l,r}$), the qubit states are dominated by different charge configurations— $(2, 0, 1)$ and $(1, 0, 2)$ —and are therefore of charge character (see Fig. 2e). Such a charge qubit has a strong electric dipole moment and is susceptible to charge noise, but also couples more strongly to resonator photons. We operate our qubit in the regime $|\Delta| \lesssim t_{l,r}$, in which we quantify the spin and charge character of the qubit states as follows: for each of the qubit states $|0_q\rangle$ and $|1_q\rangle$, we define $P_{(1,1,1)}$ to be the sum of the occupation probabilities of the three $(1, 1, 1)$ basis states, and $P_{(2,0,1)}$ and $P_{(1,0,2)}$ to be the occupation probabilities of the $(2, 0, 1)$ and $(1, 0, 2)$ states, respectively. These quantities depend on Δ , as depicted in Fig. 2e, in which t_l , t_r and ε are the same as for the measurement of the vacuum Rabi mode splitting in Fig. 1c.

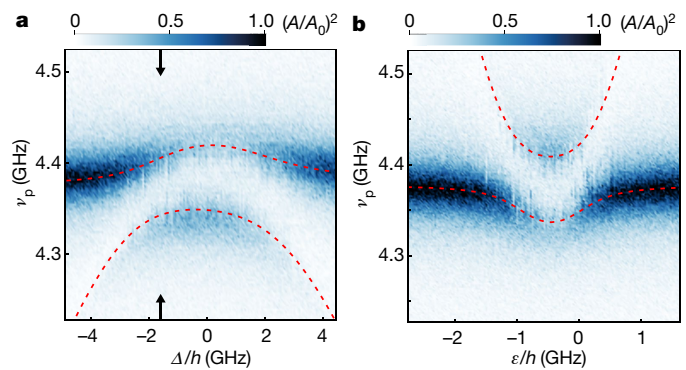


Fig. 4 | Resonator spectrum at the saddle point in the qubit energy. **a**, Resonator transmission $(A/A_0)^2$ as a function of probe frequency ν_p and detuning Δ for $\varepsilon/h = -1.03$ GHz, $t_l/h = 9.04$ GHz and $t_r/h = 7.99$ GHz. The arrows indicate the position of the vacuum Rabi mode splitting shown in Fig. 1c. **b**, Resonator transmission as a function of asymmetry ε for $\Delta/h = 0.23$ GHz, $t_l/h = 8.25$ GHz and $t_r/h = 8.64$ GHz. The dashed lines in **a** and **b** are the eigenenergies of the coupled qubit–resonator system.

The value of $\Delta/h = -1.44$ GHz used for the vacuum Rabi measurement is indicated in Fig. 2e by a vertical yellow line, at which point both qubit states have a high $P_{(1,1,1)}$. A majority of the quantum information is stored in the spin degree of freedom, providing protection from charge decoherence. On the other hand, a finite qubit–photon coupling is generated by the admixture with asymmetric charge states^{20,21}, apparent as finite $P_{(1,0,2)}$ and $P_{(2,0,1)}$ in Fig. 2e, similarly to other spin–qubit implementations^{26,27}. The amount of charge admixture and hence the nature of the qubit in our system is electrostatically tunable with the parameter Δ . This is quantified in the spin–photon coupling strength g_s , which is approximated in our qubit-operation regime as $g_s = [1/2 + \sqrt{2}/24 \times (3 + \sqrt{3})\Delta/t]g_c$, where g_c is the charge–photon coupling strength (Supplementary Information, section S2). We obtain $g_c/(2\pi) = 71$ MHz from the vacuum Rabi measurement in Fig. 1c.

Qubit–resonator interaction

Next we probe the energy spectrum of the qubit with the resonator. The theoretically expected lines of constant qubit energy as a function of detuning Δ and asymmetry ε are indicated in Fig. 3a. At constant and equal tunnel couplings, the qubit energy exhibits a saddle point at $\varepsilon = \Delta = 0$, (labelled in Fig. 3a). At this point, the qubit energy is insensitive to dephasing in the ε and Δ directions to first order²². To extract contours of qubit energy, we apply a microwave probe tone at frequency ν_p on-resonance with the resonator ($\nu_p = \nu_r$), tune the qubit energy E_q with ε and Δ , and measure the phase of the signal that is transmitted through the resonator (Fig. 3b–d). We observe a phase shift whenever the qubit and the resonator approach a resonance, $E_q = h\nu_r$. When the resonance is crossed, the phase changes sign. Determining these transition points in the ε – Δ plane experimentally at fixed tunnel couplings maps the energy contour $E_q(\Delta, \varepsilon) = h\nu_p$, reproducing one of the theoretically expected energy contours shown in Fig. 3a. We map different energy contours by changing the tunnel coupling. This is realized experimentally by changing the electrical potential of the gate lines between the plunger gates (see Fig. 1b).

From Fig. 3b to Fig. 3d, we increase the average tunnel coupling to map different contour lines of E_q (as labelled in Fig. 3a). We obtain the magnitude of both tunnel barriers for Fig. 3b–d from a fit to the resonance positions of the phase-response data. A simultaneous fit to the three datasets in Fig. 3b–d reduces the number of free parameters (Supplementary Information, section S4) and results in excellent agreement between theoretical and measured resonance conditions. The tunability of the position of the resonator–qubit resonance via the tunnel coupling allows us to observe qubit–photon coupling at the saddle point in the qubit energy in Fig. 3c. Note that, as observed in

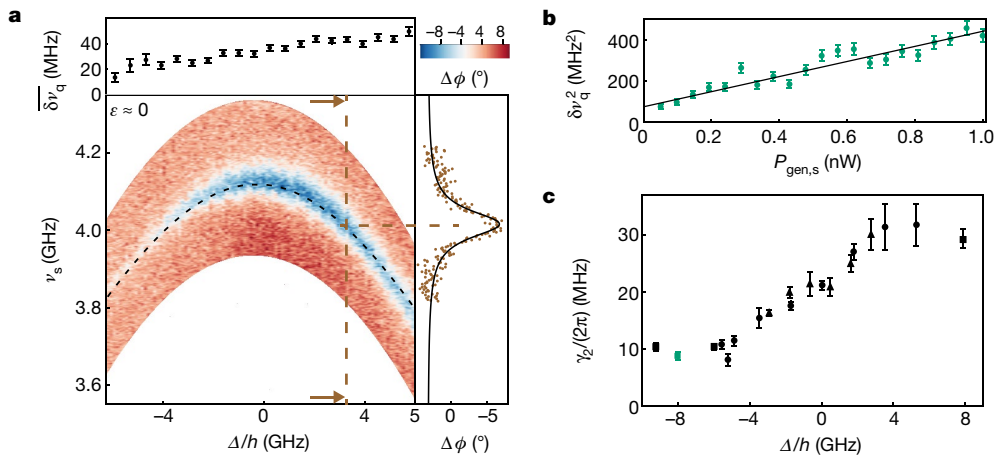


Fig. 5 | Qubit spectroscopy. **a**, Phase response of the resonator probed on-resonance as a function of spectroscopy frequency ν_s and detuning Δ , with ε set to the minimum of the qubit energy in ε ($\varepsilon \approx 0$), for $t_l/h = 8.10$ GHz, $t_r/h = 7.86$ GHz, a drive-generator power of $P_{\text{gen},s} = 0.75$ nW and a resonator photon occupation of less than 1. The theoretically expected position of the phase-response minimum is indicated by a dashed line. On the right, a Lorentzian with a half-width at half-maximum of $\delta\nu_q$ (black line) is fitted to a cut of the phase response (brown dashed line in the main panel; brown points). At the top we show $\delta\nu_q$ (points), which is the average of $\delta\nu_q$ over five subsequent cuts along Δ , along with its standard

error (error bars). **b**, Dependence of $\delta\nu_q$ (with standard errors) on the drive-generator power $P_{\text{gen},s}$, measured at $\Delta/h = -8.03$ GHz and with ε set to the minimum of the qubit energy in ε , for $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz. The solid line is a fit to the expected linear dependence. **c**, Extracted qubit decoherence $\gamma_2/(2\pi)$ (with standard errors) as a function of Δ for three different tunnel-coupling configurations: $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz (squares), $t_l/h = 7.47$ GHz and $t_r/h = 7.77$ GHz (triangles), and $t_l/h = 8.10$ GHz and $t_r/h = 7.86$ GHz (circles). The value obtained from the linear fit in **b** is shown in green.

error (error bars). **b**, Dependence of $\delta\nu_q$ (with standard errors) on the drive-generator power $P_{\text{gen},s}$, measured at $\Delta/h = -8.03$ GHz and with ε set to the minimum of the qubit energy in ε , for $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz. The solid line is a fit to the expected linear dependence. **c**, Extracted qubit decoherence $\gamma_2/(2\pi)$ (with standard errors) as a function of Δ for three different tunnel-coupling configurations: $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz (squares), $t_l/h = 7.47$ GHz and $t_r/h = 7.77$ GHz (triangles), and $t_l/h = 8.10$ GHz and $t_r/h = 7.86$ GHz (circles). The value obtained from the linear fit in **b** is shown in green.

error (error bars). **b**, Dependence of $\delta\nu_q$ (with standard errors) on the drive-generator power $P_{\text{gen},s}$, measured at $\Delta/h = -8.03$ GHz and with ε set to the minimum of the qubit energy in ε , for $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz. The solid line is a fit to the expected linear dependence. **c**, Extracted qubit decoherence $\gamma_2/(2\pi)$ (with standard errors) as a function of Δ for three different tunnel-coupling configurations: $t_l/h = 8.74$ GHz and $t_r/h = 8.12$ GHz (squares), $t_l/h = 7.47$ GHz and $t_r/h = 7.77$ GHz (triangles), and $t_l/h = 8.10$ GHz and $t_r/h = 7.86$ GHz (circles). The value obtained from the linear fit in **b** is shown in green.

Tunable qubit coherence and coupling strength

To characterize the spin qubit further, we now consider the shift in the resonator frequency due to qubit–resonator coupling in the dispersive regime, in which the qubit–resonator detuning is much larger than the qubit–photon coupling strength²⁸. In addition to the resonator probe tone at frequency $\nu_p = \nu_r$, a spectroscopy tone at frequency ν_s is applied to the right plunger gate, indicated in Fig. 1b. At resonance with the qubit ($E_q = h\nu_s$), the drive excites the qubit from its ground state $|0_q\rangle$ to the excited state $|1_q\rangle$. This results in a dispersive shift in the resonator frequency, which we detect as a drop in the phase-response signal. By sweeping both the detuning Δ and the spectroscopy frequency ν_s , with ε set to the minimum of the qubit energy in ε , we trace the spectroscopic qubit signal (Fig. 5a). This signal resembles the Δ dependence of the observed (Fig. 4a) and calculated (Fig. 3a) qubit energy and is in good agreement with theory (dashed line in Fig. 5a).

The qubit decoherence $\gamma_2/(2\pi)$ is equal to the half-width at half-maximum ($\delta\nu_q$) of the spectroscopic dip in the phase signal in the limit of

zero drive power ($P_{\text{gen},s} \rightarrow 0$)²⁸. For finite drive power, such as in Fig. 5a, the spectroscopic signal is power-broadened²⁸. We define $\delta\nu_q$ as the average of $\delta\nu_q$ over five cuts along the Δ direction in Fig. 5a and observe an increase in $\delta\nu_q$ with increasing Δ (top panel of Fig. 5a). To distinguish the effects of power broadening and qubit decoherence on $\delta\nu_q$, we extract γ_2 (Fig. 5c) by measuring $\delta\nu_q$ as a function of the power of the spectroscopy tone (Fig. 5b) for different Δ and three different tunnel-coupling configurations. We estimate the Purcell decay and the measurement-induced dephasing to be at least one order of magnitude smaller than $\gamma_2/(2\pi)$ (Fig. 5c)^{28,29}. For a high admixture of asymmetric charge states, we measure a maximum decoherence rate of $\gamma_2/(2\pi) \approx 30$ MHz. For a spin qubit with a more (1, 1, 1)-like character, we extract a minimum decoherence rate of $\gamma_2/(2\pi) \approx 10$ MHz, which corresponds to a dephasing time of $T_2^* = 1/\gamma_2 = 16$ ns. This measurement demonstrates that storing the quantum information in the spin degree of freedom increases the coherence of the qubit.

For a theoretical model that describes the data in Fig. 5c quantitatively, different sources of noise would need to be considered³⁰. Charge noise that originates from electric-field fluctuations such as gate-voltage noise leads to dephasing, which is minimal at the saddle point in the qubit energy. We observe that γ_2 is not minimal at this point ($\Delta \approx 0$ in Fig. 5c). This indicates that other noise sources, such as second-order charge-noise dephasing or phonons, are responsible for the observed qubit decoherence³¹. Another source of noise is the fluctuating Overhauser field in the GaAs host material³², which leads to inhomogeneous broadening of the line width of the qubit. This is a likely explanation for the lower limit of $\gamma_2/(2\pi) \approx 10$ MHz in Fig. 5a, consistent with previous studies that reported similar dephasing times for a resonant exchange qubit³³ and other spin qubits in GaAs^{34,35}. To distinguish and quantify the contributions of the aforementioned noise sources to the experimental qubit decoherence, additional analysis such as time-resolved measurements is necessary.

Finally, we show that the average photon number in the resonator is well below 1 for the measurement of the Rabi splitting. In the dispersive regime, the qubit frequency ν_q shifts as a function of the number of photons n in the resonator, which depends linearly on the power $P_{\text{gen},r}$ at the generator of the resonator probe tone. In addition, there is a Lamb shift in the qubit frequency due to the coupling to vacuum fluctuations. This results in a dressed qubit frequency $\tilde{\nu}_q = \nu_q + (2n + 1)[g_s/(2\pi)]^2/(\nu_q - \nu_r)$ (ref. 28). In Fig. 6a, we observe

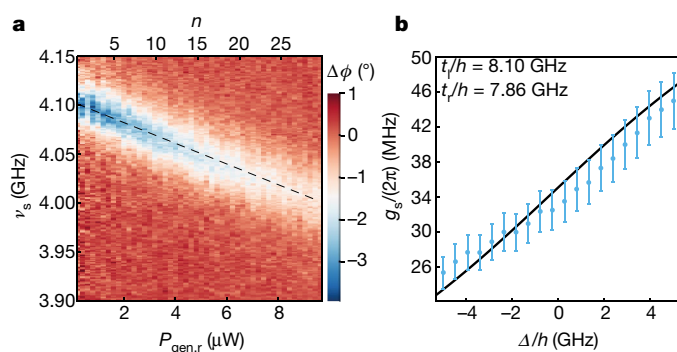


Fig. 6 | AC Stark shift. **a**, Phase response as a function of spectroscopy frequency ν_s and the power $P_{\text{gen},r}$ at the generator of the resonator probe tone. $P_{\text{gen},r}$ is converted to the average number of photons in the resonator n . The drive-generator power is set to $P_{\text{gen},s} = 0.25$ nW and the resonator is probed on-resonance. The qubit parameters are tunnel couplings $t_l/h = 8.72$ GHz and $t_r/h = 8.18$ GHz, detuning $\Delta/h = -6.0$ GHz and asymmetry $\varepsilon/h = -0.26$ GHz. The minimum in the phase response is indicated by a dashed line. **b**, Spin-qubit-photon coupling strength g_s , with errors from the calibration of the photon number, as a function of Δ (points), compared to the prediction from theory (line) for ε close to the minimum of the qubit energy in ε .

the frequency shift due to the AC Stark shift in the spectroscopic qubit signal measured at $\Delta/h = -6.02$ GHz and $\varepsilon/h = -0.26$ GHz. At this operating point, we obtain g_s from an independent measurement of the shift in the resonator frequency, similar to the one displayed in Fig. 4b (Supplementary Information, section S5). From a linear fit to the power-dependent dressed qubit frequency in Fig. 6a, we obtain the calibration factor $\alpha = n/P_{\text{gen},r} \approx 3 \times 10^{-3}$ photons nW $^{-1}$. The vacuum Rabi splitting shown in Fig. 1c was recorded for $P_{\text{gen},r} = 100$ nW. We can therefore reliably claim that for this measurement the average number of photons in the resonator is roughly 0.3. This confirms that we indeed achieved strong hybridization of the spin qubit with a single microwave photon.

With the known calibration factor α , the AC Stark shift provides direct access to the qubit-photon coupling strength (Supplementary Information, section S5). We observe in Fig. 6b that the coupling strength increases with increasing Δ . Because the contribution of the (1, 0, 2) and (2, 0, 1) charge configurations to the qubit states increases with Δ , the electric dipole moment of the qubit states and hence the qubit-resonator coupling is enhanced. However, this increase in coupling strength comes at the cost of an increase in qubit decoherence (see Fig. 5c). Our theoretical model describes this behaviour quantitatively.

Conclusion

We have coherently coupled a resonant exchange qubit to single microwave photons in a circuit quantum electrodynamics architecture. The triple-quantum-dot spin qubit arises from the exchange interaction, which couples spin and charge independent of the host material. Other spin-qubit implementations have been restricted to materials with strong spin-orbit coupling¹⁰ or require additional components such as ferromagnets^{11,26,27} for spin-charge hybridization. Furthermore, the triple-quantum-dot spin qubit is versatile because all of its parameters can be controlled electrostatically. For these reasons, it is possible to move our architecture to material systems with minimal hyperfine interaction, such as graphene³⁶ or isotopically purified silicon⁸, without the need to deposit ferromagnetic materials, which is generally undesirable in the presence of a superconductor. By doing so, we expect the qubit coherence to improve by at least one order of magnitude.

While writing up our results we became aware of independent but related work that demonstrates strong spin-photon coupling in a double-quantum-dot spin qubit in silicon^{26,27}.

Data availability

The data related to this study are available from the corresponding author on reasonable request.

Received: 4 January 2018; Accepted: 19 June 2018;

Published online 25 July 2018.

- DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).
- Majer, J. et al. Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
- Wallraff, A. et al. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Bruhat, L. E. et al. Strong coupling between an electron in a quantum dot circuit and a photon in a cavity. Preprint at <https://arxiv.org/abs/1612.05214> (2016).
- Mi, X., Cady, J. V., Zajac, D. M., Deelman, P. W. & Petta, J. R. Strong coupling of a single electron in silicon to a microwave photon. *Science* **355**, 156–158 (2017).
- Stockklauser, A. et al. Strong coupling cavity QED with gate-defined double quantum dots enabled by a high impedance resonator. *Phys. Rev. X* **7**, 011030 (2017).
- Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. K. Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79**, 1217–1265 (2007).
- Zwanenburg, F. A. et al. Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013).
- Schoelkopf, R. J. & Girvin, S. M. Wiring up quantum systems. *Nature* **451**, 664–669 (2008).
- Petersson, K. D. et al. Circuit quantum electrodynamics with a spin qubit. *Nature* **490**, 380–383 (2012).
- Viennot, J. J., Dartailh, M. C., Cottet, A. & Kontos, T. Coherent coupling of a single spin to microwave cavity photons. *Science* **349**, 408–411 (2015).
- Pioro-Ladrière, M. et al. Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nat. Phys.* **4**, 776–779 (2008).
- Hu, X., Liu, Y.-x. & Nori, F. Strong coupling of a spin qubit to a superconducting stripline cavity. *Phys. Rev. B* **86**, 035314 (2012).
- Beaudoin, F., Lachance-Quirion, D., Coish, W. A. & Pioro-Ladrière, M. Coupling a single electron spin to a microwave resonator: controlling transverse and longitudinal couplings. *Nanotechnology* **27**, 464003 (2016).
- Medford, J. et al. Quantum-dot-based resonant exchange qubit. *Phys. Rev. Lett.* **111**, 050501 (2013).
- Medford, J. et al. Self-consistent measurement and state tomography of an exchange-only spin qubit. *Nat. Nanotechnol.* **8**, 654–659 (2013).
- Gaudreau, L. et al. Coherent control of three-spin states in a triple quantum dot. *Nat. Phys.* **8**, 54–58 (2012).
- Taylor, J. M., Srinivasa, V. & Medford, J. Electrically protected resonant exchange qubits in triple quantum dots. *Phys. Rev. Lett.* **111**, 050502 (2013).
- Russ, M. & Burkard, G. Asymmetric resonant exchange qubit under the influence of electrical noise. *Phys. Rev. B* **91**, 235411 (2015).
- Srinivasa, V., Taylor, J. M. & Tahan, C. Entangling distant resonant exchange qubits via circuit quantum electrodynamics. *Phys. Rev. B* **94**, 205421 (2016).
- Russ, M. & Burkard, G. Long distance coupling of resonant exchange qubits. *Phys. Rev. B* **92**, 205412 (2015).
- Russ, M., Ginzler, F. & Burkard, G. Coupling of three-spin qubits to their electric environment. *Phys. Rev. B* **94**, 165411 (2016).
- Devoret, M., Girvin, S. & Schoelkopf, R. Circuit-QED: how strong can the coupling between a Josephson junction atom and a transmission line resonator be? *Ann. Phys.* **16**, 767–779 (2007).
- Samkharadze, N. et al. High-kinetic-inductance superconducting nanowire resonators for circuit QED in a magnetic field. *Phys. Rev. Appl.* **5**, 044004 (2016).
- Collett, M. J. & Gardiner, C. W. Squeezing of intracavity and traveling-wave light fields produced in parametric amplification. *Phys. Rev. A* **30**, 1386–1391 (1984).
- Mi, X. et al. A coherent spin-photon interface in silicon. *Nature* **555**, 599–603 (2018).
- Samkharadze, N. et al. Strong spin-photon coupling in silicon. *Science* **359**, 1123–1127 (2018).
- Schuster, D. I. et al. ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
- Sete, E. A., Gambetta, J. M. & Korotkov, A. N. Purcell effect with microwave drive: suppression of qubit relaxation rate. *Phys. Rev. B* **89**, 104516 (2014).
- Russ, M. & Burkard, G. Three-electron spin qubits. *J. Phys. Condens. Matter* **29**, 393001 (2017).
- Mehl, S. & DiVincenzo, D. P. Noise analysis of qubits implemented in triple quantum dot systems in a Davies master equation approach. *Phys. Rev. B* **87**, 195309 (2013).
- Hung, J.-T., Fei, J., Friesen, M. & Hu, X. Decoherence of an exchange qubit by hyperfine interaction. *Phys. Rev. B* **90**, 045308 (2014).
- Malinowski, F. K. et al. Symmetric operation of the resonant exchange qubit. *Phys. Rev. B* **96**, 045443 (2017).
- Petta, J. R. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
- Koppens, F. H. L., Nowack, K. C. & Vandersypen, L. M. K. Spin echo of a single electron spin in a quantum dot. *Phys. Rev. Lett.* **100**, 236802 (2008).
- Trauzettel, B., Bulaev, D. V., Loss, D. & Burkard, G. Spin qubits in graphene quantum dots. *Nat. Phys.* **3**, 192–196 (2007).

Acknowledgements We acknowledge discussions with M. Russ and A. Stockklauser. We thank M. Collodo, P. Kurpiers and P. Märki for contributions to our experimental set-up. This work was supported by the Swiss National Science Foundation through the National Center of Competence in Research (NCCR) Quantum Science and Technology. U.C.M. and A.B. were supported by NSERC and the Canada First Research Excellence fund.

Reviewer information *Nature* thanks T. Meunier and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.J.L., J.V.K. and P.S. fabricated the device. A.J.L. and J.V.K. performed the experiments. A.J.L. and J.V.K. analysed the data. U.C.M. and A.B. provided theory support for the experiment. C.R. and

W.W. grew the wafer material. A.J.L., J.V.K., T.I. and U.C.M. wrote the manuscript with the input of all authors. A.W., K.E. and T.I. supervised the experiment.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0365-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.J.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Functional aspects of meningeal lymphatics in ageing and Alzheimer's disease

Sandro Da Mesquita^{1,2,10*}, Antoine Louveau^{1,2,10}, Andrea Vaccari^{3,4}, Igor Smirnov^{1,2}, R. Chase Cornelison⁴, Kathryn M. Kingsmore⁴, Christian Contarino^{1,2,5}, Suna Onengut-Gumuscu⁶, Emily Farber⁶, Daniel Raper^{1,2,7}, Kenneth E. Viar^{1,2}, Romie D. Powell^{1,2}, Wendy Baker^{1,2}, Nisha Dabhi^{1,2}, Robin Bai^{1,2}, Rui Cao⁴, Song Hu⁴, Stephen S. Rich⁶, Jennifer M. Munson^{4,8}, M. Beatriz Lopes⁹, Christopher C. Overall^{1,2}, Scott T. Acton^{3,4} & Jonathan Kipnis^{1,2*}

Ageing is a major risk factor for many neurological pathologies, but its mechanisms remain unclear. Unlike other tissues, the parenchyma of the central nervous system (CNS) lacks lymphatic vasculature and waste products are removed partly through a paravascular route. (Re)discovery and characterization of meningeal lymphatic vessels has prompted an assessment of their role in waste clearance from the CNS. Here we show that meningeal lymphatic vessels drain macromolecules from the CNS (cerebrospinal and interstitial fluids) into the cervical lymph nodes in mice. Impairment of meningeal lymphatic function slows paravascular influx of macromolecules into the brain and efflux of macromolecules from the interstitial fluid, and induces cognitive impairment in mice. Treatment of aged mice with vascular endothelial growth factor C enhances meningeal lymphatic drainage of macromolecules from the cerebrospinal fluid, improving brain perfusion and learning and memory performance. Disruption of meningeal lymphatic vessels in transgenic mouse models of Alzheimer's disease promotes amyloid- β deposition in the meninges, which resembles human meningeal pathology, and aggravates parenchymal amyloid- β accumulation. Meningeal lymphatic dysfunction may be an aggravating factor in Alzheimer's disease pathology and in age-associated cognitive decline. Thus, augmentation of meningeal lymphatic function might be a promising therapeutic target for preventing or delaying age-associated neurological diseases.

For decades, the CNS has been seen as an immune privileged organ¹, because of its limited interactions with the immune system, especially under homeostatic, healthy conditions^{2,3}. Immune cells do not enter the parenchyma of the healthy brain as such; the surveillance of the CNS takes place within the meningeal spaces, where a great variety of immune cells is found^{2,3}. Our group, along with others^{4,5}, has recently (re)discovered and characterized the lymphatic vessels within the meninges (of rodents⁴, non-human primates and humans⁶), although the role of these vessels in CNS function and in pathologies remains unclear.

Body tissues are perfused by interstitial fluid (ISF), which is locally reabsorbed via the lymphatic vascular network. By contrast, the parenchyma of the CNS is devoid of lymphatic vasculature²; in the brain, removal of cellular debris and toxic molecules, such as amyloid- β peptides, is mediated by a combination of transcellular transport mechanisms across the blood–brain and blood–cerebrospinal fluid (CSF) barriers^{7–9}, phagocytosis and digestion by resident microglia and recruited monocytes and/or macrophages^{10,11}, as well as CSF influx and ISF efflux through a paravascular (glymphatic) route^{12–14}. The (re)discovery and characterization of meningeal lymphatic vessels has led to a reassessment of the pathways for the clearance of waste from the CNS^{4,5}. The role of this vasculature in brain function, specifically in the context of ageing and Alzheimer's disease, has not been studied. Alzheimer's disease is the most common form of dementia and its prevalence increases with age^{15,16}. Extracellular deposition of amyloid- β aggregates, the main constituent of senile plaques, is considered to be a pathological hallmark of Alzheimer's disease that contributes to neuronal dysfunction and behavioural changes^{16,17}. It is interesting to note that the amyloid- β protein was initially isolated from homogenates of

meningeal tissue from patients with Alzheimer's disease¹⁸. However, the mechanisms that underlie the accumulation of amyloid- β in the brain and meninges of patients with Alzheimer's disease are still not fully understood. The ageing-associated decrease in paravascular recirculation of CSF and ISF¹³ is thought to be responsible, at least in part, for the accumulation of amyloid- β in the brain parenchyma^{12,13,19}. Ageing also leads to progressive lymphatic vessel dysfunction in peripheral tissues^{20–22}. However, little is known about a possible functional decay of the CSF-draining meningeal lymphatic vessels with age and how this decay might influence CNS amyloid- β pathology in Alzheimer's disease.

Here we show that meningeal lymphatic vessels have an essential role in maintaining brain homeostasis by draining macromolecules from the CNS (both CSF and ISF) into the cervical lymph nodes. Using pharmacological, surgical and genetic models, we show that impairment or enhancement of meningeal lymphatic function in mice affects paravascular influx of CSF macromolecules, efflux of ISF macromolecules and cognitive task performance. Our findings demonstrate that meningeal lymphatic vessel dysfunction may be one of the underlying factors for worsened amyloid- β pathology and cognitive deficits in Alzheimer's disease and might be therapeutically targeted to alleviate age-associated cognitive decline.

Meningeal lymphatics and brain perfusion

Given the close communication and continuous exchange of molecular contents between the CSF and ISF^{5,12}, we hypothesized that brain influx of CSF macromolecules through the paravascular pathway is affected by the meningeal lymphatic vessels. To test this hypothesis, we

¹Center for Brain Immunology and Glia (BIG), University of Virginia, Charlottesville, VA, USA. ²Department of Neuroscience, University of Virginia, Charlottesville, VA, USA. ³Virginia Image and Video Analysis Laboratory, Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, USA. ⁴Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA. ⁵Department of Mathematics, University of Trento, Povo, Italy. ⁶Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. ⁷Department of Neurosurgery, University of Virginia Health System, Charlottesville, VA, USA. ⁸Department of Biomedical Engineering and Mechanics, College of Engineering, Virginia Tech, Blacksburg, VA, USA. ⁹Department of Pathology, University of Virginia, Charlottesville, VA, USA. ¹⁰These authors contributed equally: Sandro Da Mesquita, Antoine Louveau. *e-mail: sd8tf@virginia.edu; kipnis@virginia.edu

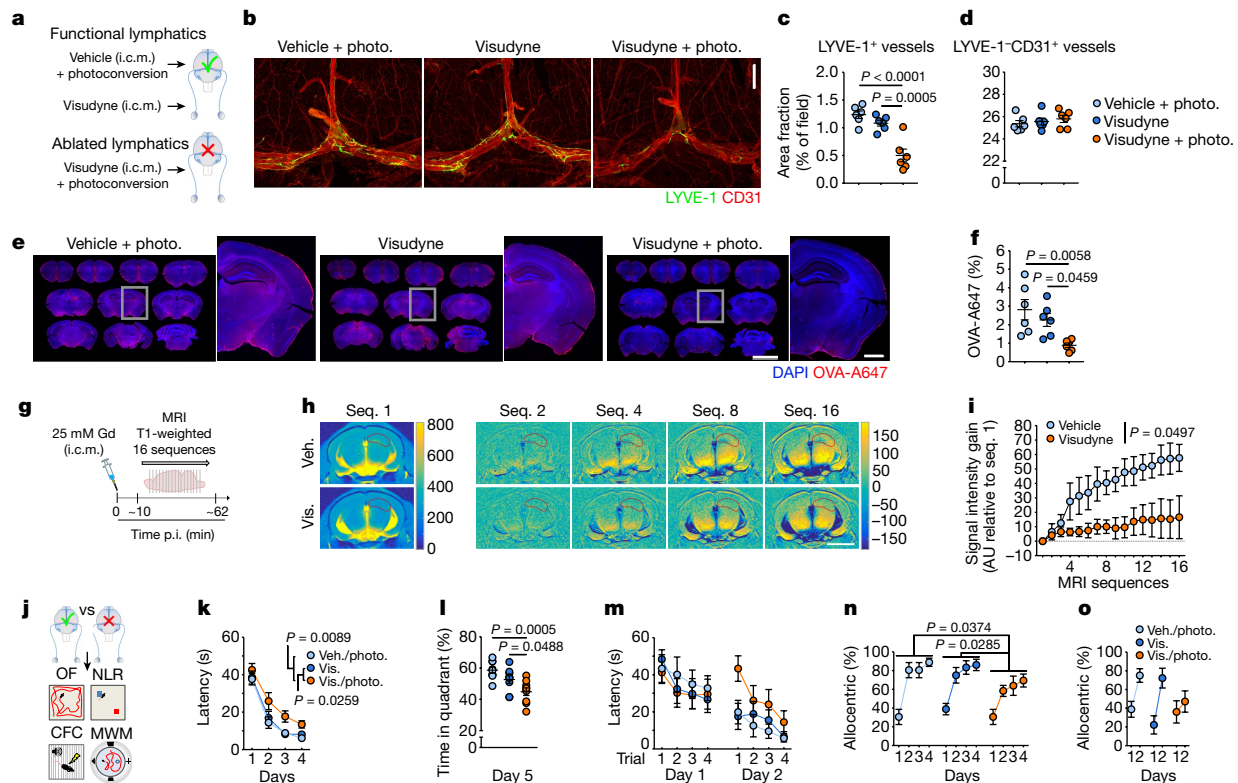


Fig. 1 | Impairing meningeal lymphatics affects brain CSF influx and ISF diffusion and worsens cognitive function. **a**, Seven days after lymphatic ablation, mice were injected with 5 μ l of OVA-A647 i.c.m. **b**, Representative images of meningeal whole-mounts stained for LYVE-1 and CD31. Scale bar, 1 mm. **c**, Quantification of area fraction (%) occupied by LYVE-1⁺ lymphatic vessels (**c**) and LYVE-1[−]CD31⁺ blood vessels (**d**). **e**, Representative brain sections showing 4',6-diamidino-2-phenylindole (DAPI) and OVA-A647. Scale bars, 5 mm and 1 mm (inset). **f**, Quantification of OVA-A647 area fraction. Data in **c**, **d** and **f** are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test. **a–f**, Data are representative of two independent experiments; significant differences between vehicle (Veh.) with photoconversion (photo.) and visudyne (Vis.) with photoconversion were replicated in five independent experiments. **g**, Gd was injected (i.c.m.) and T1-weighted MRI acquisition was performed seven days after meningeal lymphatic ablation. p.i., post injection. **h**, Representative images of sequence (Seq.) 1 and of Gd intensity gain in subsequent sequences. The hippocampus

is delineated in red. Scale bar, 3 mm. **i**, Quantification of the Gd signal intensity gain over 16 sequences (relative to sequence 1) in hippocampus. Data in **i** are mean \pm s.e.m., $n = 4$ per group, repeated-measures two-way ANOVA with Bonferroni's post hoc test. **g–i** Data are representative of two independent experiments. **j**, Meningeal lymphatic ablation was performed twice and two weeks after the last intervention, open field (OF), NLR, CFC and MWM behavioural tests were performed (see Extended Data Fig. 5 for open field, NLR and CFC tests). **k**, Latency to platform (acquisition). **l**, Percentage of time spent in the target quadrant (probe). **m**, Latency to platform (reversal). **n**, **o**, Allocentric navigation strategies (%) used in the MWM acquisition (**n**) and reversal (**o**). Data in **k–o** are mean \pm s.e.m., $n = 9$ per group; repeated-measures two-way ANOVA with Bonferroni's post hoc test (**k**, **m–o**), one-way ANOVA with Bonferroni's post hoc test (**l**); significant differences between vehicle with photoconversion and visudyne with photoconversion were replicated in three independent experiments.

ablated meningeal lymphatic vessels by injecting a photodynamic drug, visudyne (also known as verteporfin for injection), into the CSF, which upon photoconversion has been shown to preferentially damage the lymphatic endothelial cells (LECs)^{23,24}. Injections of vehicle followed by photoconversion and injections of visudyne without the photoconversion step were used as two controls (Fig. 1a). The use of this method resulted in effective ablation of meningeal lymphatic vessels (Fig. 1b, c), without any detectable off-target effects in the coverage of meningeal blood vasculature seven days after the procedure (Fig. 1d). To confirm functional impairment after meningeal lymphatic vessel ablation, we injected 5 μ l of fluorescent ovalbumin–Alexa Fluor 647 (OVA-A647; approximately 45 kDa) into the cisterna magna (i.c.m.) and measured the drainage of this tracer from the CSF into the deep cervical lymph nodes (dCLNs) (Extended Data Fig. 1a). A significant reduction in OVA-A647 drainage was observed in the visudyne with photoconversion group compared to the control groups (Extended Data Fig. 1b). Notably, the structure of major intracranial veins and arteries was not altered (Extended Data Fig. 1c–h). Similarly, the integrity of the blood–brain barrier, assessed by T1-weighted magnetic resonance imaging (MRI) after intravenous injection of gadolinium (Gd) as contrast agent (Extended Data Fig. 1i, j), and the ventricular volume measured by T2-weighted SPACE (sampling perfection with application optimized

contrasts using different flip angle evolution) MRI (Extended Data Fig. 1k–m) also remained unaltered after ablation of meningeal lymphatic vessels.

To avoid any confounding effects due to increased intracranial pressure (ICP) after i.c.m. injection, we measured changes in ICP after injecting different volumes of OVA-A647 (Extended Data Fig. 2a, b). There was a transient increase in ICP during i.c.m. injection of the tracer, followed by a drop in ICP upon removal of the syringe after the injection (Extended Data Fig. 2a). Mice injected with 2 μ l presented ICP values lower than baseline even 120 min post-injection (Extended Data Fig. 2b). Notably, ablation of meningeal lymphatic vessels led to an equal decrease in drainage to the dCLNs in mice upon injection of 2 μ l (Extended Data Fig. 2c–e) or 5 μ l of the tracer (Extended Data Fig. 1a, b).

Brain perfusion by the CSF tracer was found to be significantly lower in the visudyne with photoconversion group than in the control groups (Fig. 1e, f and Extended Data Fig. 2f, g). Similar findings for brain perfusion by CSF were observed when meningeal lymphatic drainage was disrupted by surgical ligation of the vessels afferent to the dCLNs (Extended Data Fig. 3a–d). Prospero homeobox protein 1 heterozygous (*Prox1*^{+/−}) mice, a genetic model of lymphatic vessel malfunction²⁵, also presented impaired perfusion through the brain parenchyma and impaired CSF drainage (Extended Data Fig. 3e–i). Together, three

different models of impaired meningeal lymphatic function (pharmacological, surgical and genetic) showed a significant impact on brain perfusion by CSF macromolecules.

To evaluate the effect of meningeal lymphatic ablation on the rate of brain perfusion by CSF, we injected Gd (i.c.m.) and performed brain T1-weighted MRI. Three different concentrations of Gd—1, 10 and 25 mM—were tested (Extended Data Fig. 3j, k) and, owing to a better signal-to-noise ratio, the concentration of 25 mM was used in subsequent experiments (Fig. 1g). A software package developed in-house, Lymph4D (see Supplementary Methods for more details), was used to process and analyse the images acquired by MRI. After 16 sequences of MRI acquisition (around 52 min), the observed signal gain in two brain regions (hippocampus and cortex) was significantly lower in the visudyne group compared to vehicle group (Fig. 1h, i and Extended Data Fig. 3l, m). Notably, along with the lower influx of Gd into the parenchyma, we observed higher contrast in signal intensity (over approximately 52 min) in the ventricles of visudyne-treated mice, suggesting that Gd accumulation in the CSF occurred (Extended Data Fig. 3n). Whether this observation is concomitant with ventricular CSF reflux (a phenomenon reported in patients with idiopathic normal-pressure hydrocephalus²⁶) warrants further investigation. Moreover, using the advection–diffusion model in Lymph4D, we found that mice had lower coefficient values of isotropic diffusion of Gd in the brain after meningeal lymphatic ablation (Extended Data Fig. 3o, p), suggesting that there is a lower rate of molecular diffusion in the brain parenchyma when meningeal lymphatic drainage is reduced.

Within the brain parenchyma, it was shown that aquaporin 4 (AQP4) expression by astrocytes plays an important role in the modulation of paravascular CSF macromolecule influx and efflux (through the glymphatic route)^{12,13}. Deletion of *Aqp4* in transgenic mice with Alzheimer's disease also resulted in increased amyloid- β plaque burden and exacerbated cognitive impairment¹⁹. Moreover, decreased perivascular AQP4 localization was observed in brain tissue from patients with Alzheimer's disease²⁷. We did not detect changes either in overall brain coverage by AQP4 (Extended Data Fig. 3q, r) or in perivascular localization of AQP4⁺ astrocytic endfeet between vehicle-treated and visudyne-treated mice (Extended Data Fig. 3s–v), suggesting that upon meningeal lymphatic ablation, impairment of brain perfusion by CSF is independent of AQP4.

Next, we examined whether the efflux of ISF macromolecules from the brain parenchyma would also be affected by meningeal lymphatic vessels. We used three different tracers, the smaller peptides amyloid- β_{42} -HyLite647 (approximately 4 kDa) and OVA-A647, and the large protein complex, low-density lipoprotein–BODIPY FL (LDL–BODIPY FL, around 500 kDa). One hour after stereotaxic injection, the levels of the remaining tracers were assessed in the parenchyma of mice in which the lymphatic vessels were ablated and in mice from the control groups (Extended Data Fig. 4a–h). Independently of the nature of the fluorescent tracer, higher levels of remnants were detected in the brains of mice from the visudyne with photoconversion groups compared to both control groups (Extended Data Fig. 4a–h). These findings, as has been suggested previously⁵, demonstrate that the efflux of parenchymal and/or ISF macromolecules and the drainage of these macromolecules into dCLNs are impaired as a consequence of meningeal lymphatic ablation, thus functionally connecting meningeal lymphatics with CSF influx and ISF efflux mechanisms.

To understand the implications of impaired meningeal lymphatic drainage for brain function, we performed meningeal lymphatic ablation twice, allowing a two-week interval between procedures to ensure prolonged lymphatic ablation, and then assessed the behaviour of mice in the open field, novel location recognition (NLR), contextual fear conditioning (CFC) and Morris water maze (MWM) tests (Fig. 1j). No differences between the groups were detected in total distance travelled and time spent in the centre of the arena in the open field test (Extended Data Fig. 5a, b) or in time spent with the object placed in a novel location in the NLR test (Extended Data Fig. 5c, d). A significant difference between control groups and visudyne with photoconversion group

was observed in the cued test of the CFC (Extended Data Fig. 5e, f), which points to an impairment in fear memory and in hippocampal–amygdala neuronal circuitry²⁸ in mice with impaired meningeal lymphatic vessel function. Mice with ablated meningeal lymphatic vessels also showed significant deficits in spatial learning in the MWM (Fig. 1k–o). Similar impairments in spatial learning and memory were observed in mice that had undergone lymphatic ligation (Extended Data Fig. 5g–j), supporting the notion that the observed effect is a result of dysfunctional meningeal lymphatic drainage and not an artefact of the ablation method using visudyne.

Using RNA sequencing (RNA-seq), we assessed the effect of visudyne treatment with photoconversion on hippocampal gene expression before and after the MWM. Principal component analysis showed that four weeks of meningeal lymphatic ablation did not induce significant changes in the hippocampal transcriptome (Extended Data Fig. 5k, l). However, significant differences in hippocampal gene expression were found in response to MWM performance after prolonged meningeal lymphatic ablation (Extended Data Fig. 5m, n). Contrary to what was observed without MWM performance (Extended Data Fig. 5k, l), individual samples from each group clustered together after the mice performed the test (Extended Data Fig. 5m, n). Notably, although the fold change in significantly altered genes after lymphatic ablation and MWM was moderate ($-1.79 < \log_2(\text{fold change}) < 1.69$), functional enrichment analysis (Extended Data Fig. 5o, p) revealed changes in gene sets associated with neurodegenerative diseases, such as Huntington's, Parkinson's and Alzheimer's disease (Extended Data Fig. 5o). Significant transcriptional alterations were also associated with excitatory synaptic remodelling and plasticity, hippocampal neuronal transmission²⁹, learning and memory and ageing-related cognitive decline³⁰ (Extended Data Fig. 5q, r). Furthermore, different gene sets that are involved in the regulation of metabolite generation and processing, glycolysis and mitochondrial respiration and oxidative stress were also significantly altered in the hippocampus upon lymphatic ablation and performance of the behaviour test (Extended Data Fig. 5p, s–v).

Meningeal lymphatic vessels during ageing

Ageing is the principal risk factor for many neurological disorders, including Alzheimer's disease^{15,16}, and has a detrimental effect on CSF and ISF paravascular recirculation within the brain¹³. The reported findings that ageing is also associated with peripheral lymphatic dysfunction^{20–22} led us to hypothesize that the deterioration of meningeal lymphatic vessels underlies some aspects of age-associated cognitive decline. Indeed, and in agreement with a previous study¹³, old mice demonstrate reduced brain perfusion by CSF macromolecules compared to young counterparts (Extended Data Fig. 6a, b). Impaired brain perfusion by CSF in old mice was accompanied by a decrease in meningeal lymphatic vessel diameter and coverage, as well as decreased drainage of CSF macromolecules into dCLNs in both females and males (Extended Data Fig. 6c–f). To further address the effect of ageing on meningeal lymphatic vessels, we performed RNA-seq analysis of LECs sorted from the meninges of young-adult (2–3 months of age) and old (20–24 months of age) mice (Fig. 2a–d and Extended Data Fig. 6g). Differential expression of 607 genes was detected in the meningeal LECs of old compared to young-adult mice (Fig. 2a). Of note, the expression of genes that encode classical markers of LECs, including *Flt4*, which encodes the vascular endothelial growth factor C (VEGF-C) receptor tyrosine kinase VEGFR3, was not significantly altered at 20–24 months (Fig. 2b). Enrichment analysis revealed, however, changes in gene sets involved in immune and inflammatory responses, phospholipid metabolism, extracellular matrix organization, cellular adhesion and endothelial tube morphogenesis, all of which suggest that there are functional alterations in meningeal LECs with age (Fig. 2c). The altered expression of genes involved in the transmembrane receptor protein tyrosine kinase signalling pathway in old mice, namely the downregulation of *Cdk5r1*³¹, *Adamts3*³² and *Fgfr3*³³, indicated possible changes in signalling by lymphangiogenic growth factors in old meningeal LECs (Fig. 2d).

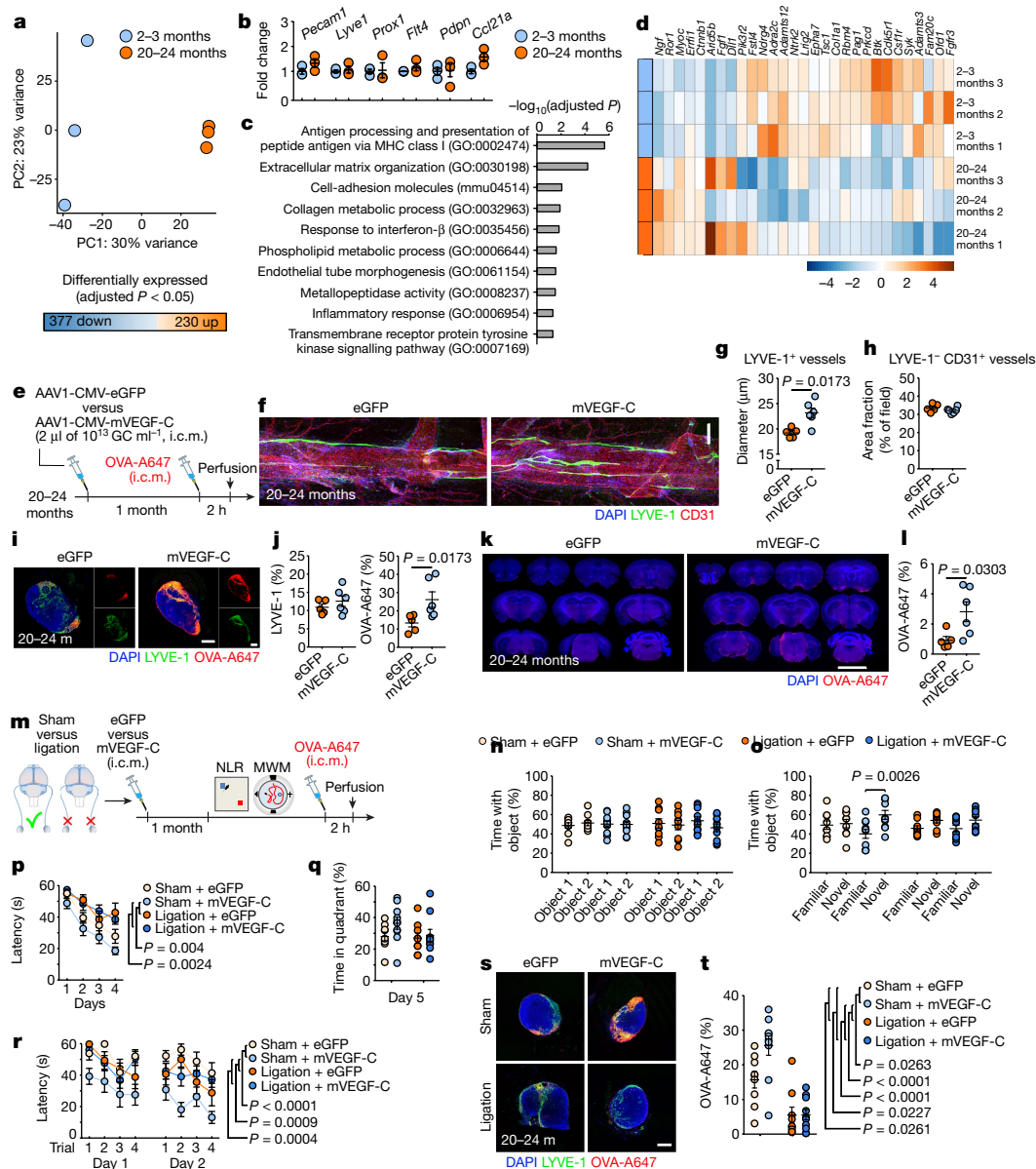


Fig. 2 | Improving meningeal lymphatic function in aged mice increases brain perfusion and alleviates cognitive deficits. **a**, Principal component (PC) analysis plot for RNA-seq of LECs from meninges of young-adult and aged mice. There were 230 genes up- and 377 genes downregulated in meningeal LECs at 20–24 months (**m**). **b**, Expression of *Pecam1*, *Lyve1*, *Prox1*, *Flt4*, *Pdpn* and *Ccl21a*. **c**, Gene sets obtained by functional enrichment of differentially expressed genes in meningeal LECs at 20–24 months. **d**, Heat map showing relative expression level of genes involved in the transmembrane receptor protein tyrosine kinase signalling pathway. Colour scale bar values represent standardized log-transformed values across samples. Data in **a–d** consist of $n = 3$ per group (individual RNA samples result from LECs pooled from 10 meninges over two independent experiments); data in **b** are mean \pm s.e.m. with two-way ANOVA with Bonferroni's post hoc test; in **a–c** P values were corrected for multiple hypothesis testing with the Benjamini–Hochberg false-discovery rate procedure; in **c**, **d** functional enrichment of differential expressed genes was performed using gene sets from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) and determined with Fisher's exact test. **e**, Old mice were injected (i.c.m.) with 2 μ l of AAV1-CMV-eGFP (eGFP) or AAV1-CMV-mVEGF-C (mVEGF-C), at 10^{13} genome copies (GC) ml^{-1} . One month later, OVA-A647 was injected i.c.m. **f**, Insets of the superior sagittal sinus showing DAPI, LYVE-1 and CD31 staining. Scale bar, 200 μ m. **g**, **h**, Quantification of the diameter of LYVE-1⁺

lymphatic vessels (**g**) and the area fraction (%) of LYVE-1⁺ CD31⁺ blood vessels (**h**). **i**, Representative sections of dCLNs showing DAPI, LYVE-1 and OVA-A647 staining. Scale bar, 200 μ m. **j**, Quantification of LYVE-1 and OVA-A647 area fraction in dCLNs. **k**, Representative brain coronal sections showing DAPI and OVA-A647 staining. Scale bar, 5 mm. **l**, Quantification of OVA-A647 area fraction in brain sections. Data in **g**, **h**, **j**, **l** are mean \pm s.e.m., $n = 5$ mice treated with eGFP, $n = 6$ mice treated with mVEGF-C, two-tailed Mann–Whitney U -test; **e–l**, Data are representative of two independent experiments. **m**, Old mice were injected with eGFP or mVEGF-C viruses (i.c.m.) after ligation of the lymphatic vessels afferent to the dCLNs or sham surgery. One month later, learning and memory was assessed in the NLR and MWM tests and mice were injected (i.c.m.) with OVA-A647. **n**, **o**, Time with the object (%) was assessed in the training (**n**) and novel location (**o**) tasks of the NLR test. **p**, Latency to platform (acquisition). **q**, Percentage of time spent in the target quadrant (probe). **r**, Latency to platform (reversal). **s**, Representative sections of dCLNs showing DAPI, LYVE-1 and OVA-A647 staining. Scale bar, 200 μ m. **t**, Quantification of OVA-A647 area fraction in dCLNs. Data in **n–r**, **t** are mean \pm s.e.m., $n = 9$ in sham with eGFP and ligation with eGFP groups, $n = 10$ in sham with mVEGF-C and ligation with mVEGF-C groups, two-way ANOVA with Bonferroni's post hoc test (**n**, **o**, **q**, **t**), repeated-measures two-way ANOVA with Bonferroni's post hoc test (**p**, **r**); **m–t** results from two independent experiments.

We have previously shown that treatment with recombinant VEGF-C increases the diameter of meningeal lymphatic vessels⁴. Furthermore, delivery of VEGF-C by adenoviral gene therapy was previously found

to efficiently boost peripheral lymphatic sprouting and function^{34,35}. A similar adenoviral serotype 1 (AAV1) vector was used here to express mouse (m)VEGF-C or enhanced green fluorescent

protein (eGFP) as control. At two and four weeks post i.c.m. injection, AAV1-infected cells expressing eGFP were found to be limited to the pia around the brain, meninges (dura and arachnoid), and pineal gland (Extended Data Fig. 6h–j). Treatment of young mice with AAV1-CMV-mVEGF-C resulted in a significant increase in meningeal lymphatic vessel diameter, without affecting blood vessel coverage (Extended Data Fig. 6k–m).

Treatment of old mice (at 20–24 months) with AAV1-CMV-mVEGF-C also resulted in increased lymphatic vessel diameter (compared to AAV1-CMV-eGFP) without detectable off-target effects on the meningeal blood vasculature coverage and on meningeal and/or brain vascular haemodynamics (Fig. 2e–h and Extended Data Fig. 6n–p). One month after AAV1-CMV-mVEGF-C treatment, old mice showed a significant increase in CSF tracer drainage into the dCLNs, which was not due to increased lymphatic vessel coverage in the nodes (Fig. 2i, j). Notably, the rate of tracer influx into the brain parenchyma was significantly increased as a result of enhanced meningeal lymphatic function (Fig. 2k, l and Extended Data Fig. 6q, r).

Transcranial delivery (through a thinned skull surface) of hydrogel-encapsulated VEGF-C peptide also resulted in increased diameter of meningeal lymphatics in young and old mice (Extended Data Fig. 7a–c). This VEGF-C treatment led to a significant increase in the function of meningeal lymphatic vessels in old mice, whereas young–adult mice did not respond to the treatment (Extended Data Fig. 7d, e), probably due to the ceiling effect of their existing capacity to drain OVA-A647. The increased drainage after VEGF-C treatment in old mice also correlated with enhanced brain perfusion by CSF macromolecules (Extended Data Fig. 7f, g).

To avoid potential off-target effects of VEGF-C on the blood vasculature through VEGFR2, we carried out transcranial delivery of VEGF-C156S (Extended Data Fig. 7h), a mutated version of VEGF-C that binds specifically to VEGFR3 and spares its effects on VEGFR2^{34,36}. Treatment with VEGF-C156S resulted in a significant increase in meningeal lymphatic diameter (Extended Data Fig. 7i, j), drainage of tracer from the CSF (Extended Data Fig. 7k, l), and paravascular influx of tracer into the brains of old mice (Extended Data Fig. 7m, n).

To determine the functional role of enhanced meningeal lymphatics in the learning behaviour of mice at different ages, we again used viral delivery of mVEGF-C (Extended Data Fig. 7o–u). This method was selected to avoid submitting aged mice to consecutive surgeries, involving general anaesthesia and skull thinning. Treatment of young–adult mice with AAV1-CMV-mVEGF-C for 1 month did not improve spatial learning and memory (Extended Data Fig. 7p, s), suggesting that there is a ceiling effect in MWM performance at this age. However, AAV1-CMV-mVEGF-C treatment resulted in significant improvement in the latency to platform and in the percentage of allocentric navigation strategies, in the MWM reversal at 12–14 months (Extended Data Fig. 7q, t) and in the MWM acquisition and reversal at 20–22 months (Extended Data Fig. 7r, u), compared to AAV1-CMV-eGFP-treated age-matched mice.

Increased expression of VEGF-C in the adult brain has previously been shown to boost proliferation of neural stem cells in the hippocampus³⁷. Although spatial learning and memory in the MWM is not dependent on adult hippocampal neurogenesis³⁸, we examined the number of Ki-67-expressing cells in the hippocampal dentate gyrus of mice treated with eGFP or mVEGF-C viral vectors at 3, 12–14 and 20–22 months of age. No differences in cell proliferation in the dentate gyrus were observed after mVEGF-C treatment (Extended Data Fig. 7v, w).

To demonstrate that the beneficial effect of mVEGF-C treatment on cognitive behaviour was through improved drainage of meningeal lymphatic vessels, we injected old mice with the eGFP or mVEGF-C viruses and concomitantly ligated the lymphatic vessels afferent to dCLNs. Assessment of learning and memory was performed one month after the procedures (Fig. 2m). The beneficial effect of mVEGF-C treatment in mice from the sham group, which performed significantly better in the NLR (Fig. 2n, o) and MWM (Fig. 2p–r) tests, was abrogated in

mice in which the CSF-draining lymphatic vessels had been ligated. Accordingly, the drainage of CSF macromolecules into dCLNs was significantly higher in sham-operated mice treated with mVEGF-C compared to all other groups (Fig. 2s, t).

Dysfunctional lymphatic vessels in amyloid- β pathology

On the basis of previous findings concerning the role of paravascular CSF and ISF recirculation in the context of Alzheimer's disease^{12,14,19,27} and our present results on the interdependence between meningeal lymphatic function and brain perfusion by CSF, we proposed that modulating meningeal lymphatic function would impact the behaviour of and brain pathology in transgenic mice with Alzheimer's disease. The potential effect of mVEGF-C treatment (through viral vector delivery) was first tested on J20 transgenic mice at 6–7 months of age (Extended Data Fig. 8a–n), when mice already present marked cognitive deficits and start to show amyloid- β deposition in the brain parenchyma^{39,40}. We were not able to improve the hyperactive phenotype of J20 mice in the open field or cognitive performance in the MWM (Extended Data Fig. 8a–f). Moreover, viral expression of mVEGF-C did not significantly affect the diameter of meningeal lymphatic vessels, the level of amyloid- β in the CSF, or amyloid- β deposition in the hippocampus (Extended Data Fig. 8g–n). In order to explain the lack of effect of the mVEGF-C treatment in J20 mice, we measured meningeal lymphatic drainage in J20 mice and in wild-type littermate controls. The same measurement was performed in a more aggressive transgenic mouse model of Alzheimer's disease, the 5xFAD mice, which already have amyloid- β plaques at three months of age⁴¹ (Extended Data Fig. 8o). Independently of the model, the level of CSF tracer drained into the dCLNs was comparable between transgenic mice with Alzheimer's disease and age-matched wild-type littermates (Extended Data Fig. 8p–s). Similarly, the morphology and coverage of meningeal lymphatic vessels did not differ between wild-type and 5xFAD mice at 3–4 months of age (Extended Data Fig. 8t, u). Collectively, these data point to no apparent meningeal lymphatic dysfunction in transgenic mice with Alzheimer's disease at younger ages, which might explain the inefficacy of mVEGF-C treatment.

Although age is the major risk factor for late-onset Alzheimer's disease^{15,16}, most transgenic mouse models that mimic early-onset Alzheimer's disease develop amyloid- β pathology at young age and, therefore, may be lacking the aspect of age-related lymphatic dysfunction. To this end, we induced prolonged meningeal lymphatic ablation in 5xFAD mice by repeated (every three weeks) injection and photoconversion of visudyne for a total of 1.5 months, starting at around two months of age (Fig. 3a). Taking into account the marked deposition of amyloid- β in the brain that these mice have at approximately three months of age, surprisingly, no obvious amyloid- β deposition was detected in the meninges of 5xFAD mice from the two control groups (Fig. 3b). However, 5xFAD mice with ablated meningeal lymphatic vessels demonstrated marked deposition of amyloid- β in the meninges (Fig. 3b), as well as macrophage recruitment to large amyloid- β aggregates (Fig. 3c). Photoacoustic imaging one week after lymphatic ablation showed that there were no differences in blood flow and oxygenation between 5xFAD mice from the different groups (Extended Data Fig. 9a–c). Analysis of lymphoid and myeloid cell populations in the meninges (Extended Data Fig. 9d) demonstrated a significant increase in the number of macrophages upon lymphatic ablation compared to both control groups (Extended Data Fig. 9e), which might be correlated with increased amyloid- β deposition and inflammation in the meninges. Notably, along with meningeal amyloid- β pathology, we observed an aggravation of brain amyloid- β burden in the hippocampi of 5xFAD mice with dysfunctional meningeal lymphatic vessels (Fig. 3d–g). A similar outcome was observed in J20 transgenic mice after a total of three months of meningeal lymphatic ablation (Extended Data Fig. 9f); amyloid- β aggregates had formed in the meninges (Extended Data Fig. 9g) and the amyloid- β plaque load in the hippocampi of these mice was significantly increased (Extended Data Fig. 9h–k).

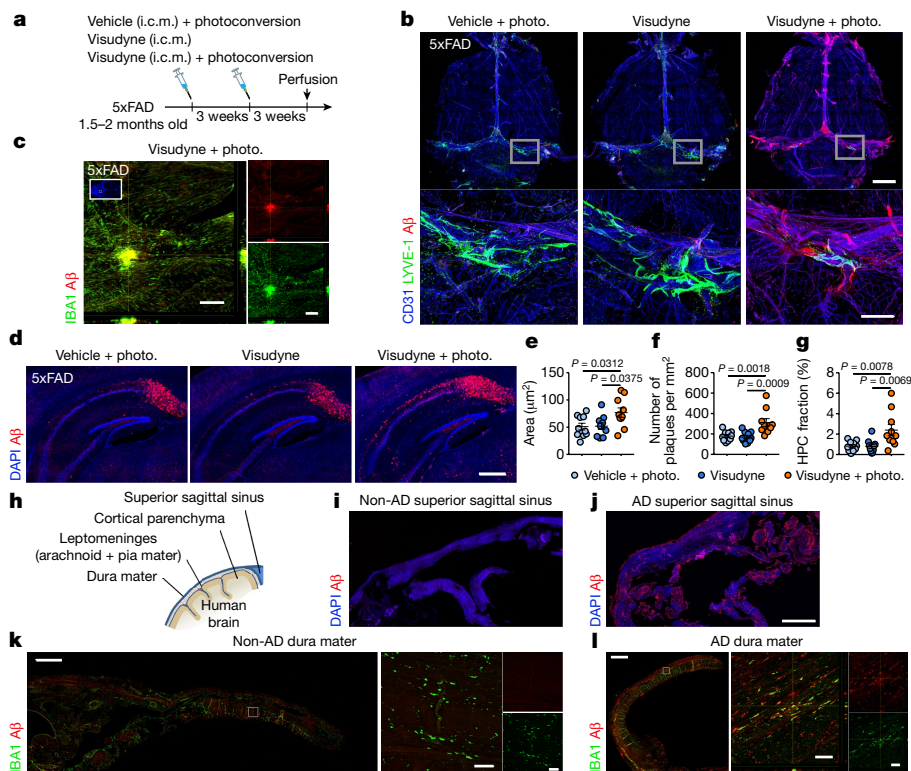


Fig. 3 | Ablation of meningeal lymphatic vessels aggravates amyloid- β pathology in transgenic mice with Alzheimer's disease. **a**, Young-adult 5xFAD mice were subjected to meningeal lymphatic ablation or control procedures. Procedures were repeated three weeks later and amyloid- β (A β) pathology was assessed six weeks after the initial intervention. **b**, Staining for CD31, LYVE-1 and amyloid- β in meninges. Scale bars, 2 mm and 500 μ m (inset). **c**, Orthogonal view of IBA $^{+}$ macrophages clustering around an amyloid- β plaque in meninges of a 5xFAD mouse with ablated lymphatic vessels. Scale bars, 200 μ m. **d**, Representative images of DAPI and amyloid- β in the hippocampus of 5xFAD mice from each group. Scale bar, 500 μ m. **e–g**, Quantification of amyloid- β plaque size (**e**), number (**f**) and coverage (**g**) in the hippocampus of 5xFAD mice. Data in **e–g** are mean \pm s.e.m., $n = 10$ per group, one-way ANOVA with

Bonferroni's post hoc test. **a–g**, Data are representative of two independent experiments. **h**, Staining for amyloid- β pathology was performed in sections of brains with attached leptomeninges (Extended Data Fig. 9) and of meningeal dura mater from patients with Alzheimer's disease and controls. **i, j**, Meningeal superior sagittal sinus tissue of a control without Alzheimer's disease (non-AD) (**i**) and a patient with Alzheimer's disease (AD) (**j**) stained with DAPI and for amyloid- β . Scale bar, 2 mm. **k, l**, Meningeal dura mater tissue of a control without Alzheimer's disease (**k**) and a patient with Alzheimer's disease (**l**) stained for IBA1 and amyloid- β . Scale bars, 1 mm and 50 μ m (orthogonal view inset). Data in **h–l** are from $n = 8$ controls and $n = 9$ patients with Alzheimer's disease and are representative of two independent experiments.

The observed meningeal amyloid- β pathology in mice after ablation of the meningeal lymphatic vessels led us to assess meningeal amyloid- β pathology in patients with Alzheimer's disease (Fig. 3h). Staining for amyloid- β in the brains of nine patients with Alzheimer's disease and eight controls without Alzheimer's disease (Extended Data Table 1) revealed, as expected, marked parenchymal deposition of amyloid- β in the brains of patients with Alzheimer's disease, but not in the brains of the controls without Alzheimer's disease (Extended Data Fig. 9l, m). Notably, when compared to tissue from controls, all samples from patients with Alzheimer's disease demonstrated striking vascular amyloid- β pathology in the cortical leptomeninges (Extended Data Fig. 9l, m) and amyloid- β deposition in the dura mater adjacent to the superior sagittal sinus (Fig. 3i, j) or further away from the sinus (Fig. 3k, l). Macrophages in the dura of cases with Alzheimer's disease were also found in close proximity to amyloid- β deposits (Fig. 3l). These findings showed that prominent meningeal amyloid- β deposition observed in patients with Alzheimer's disease is also observed in mouse models of Alzheimer's disease after meningeal lymphatic vessel ablation.

Discussion

Taken together, the present findings highlight the importance of meningeal lymphatic drainage in brain physiology. Meningeal lymphatic dysfunction in young-adult mice results in impaired brain perfusion by CSF and in learning and memory deficits. Aged mice demonstrated significant disruption of meningeal lymphatic function, which may underlie some of the aspects of age-associated cognitive

decline. Augmentation of meningeal lymphatic drainage in aged mice can ultimately facilitate the clearance of CSF and ISF macromolecules from the brain, resulting in improved cognitive function. We also show that transgenic mouse models of Alzheimer's disease recapitulate many features of brain amyloid- β pathology observed in patients with Alzheimer's disease, but not the deposition of amyloid- β observed in the dura mater. However, inducing meningeal lymphatic dysfunction in mouse models of Alzheimer's disease worsened amyloid- β pathology in the meninges and in the brain. It would be interesting to see whether transgenic mice with Alzheimer's disease, particularly the ones with a less aggressive phenotype, when sufficiently aged, would exhibit meningeal amyloid- β pathology. Furthermore, taking into account the role of the brain vascular endothelium and of other components of the blood–brain barrier, such as pericytes, in the excretion of amyloid- β from the brain^{7–9,42}, it would be very interesting to explore a possible connection between age-associated meningeal lymphatic dysfunction, impaired CSF and ISF recirculation, and decreased fitness of the blood–brain barrier and its cellular components.

Finally, it is vital to determine whether ageing-related changes in meningeal lymphatic drainage might affect the efficacy of current therapies for Alzheimer's disease, such as antibody-based treatments⁴³. Modulation of meningeal lymphatic function in aged individuals might represent a novel preventive therapeutic strategy, not only to delay initiation and progression of Alzheimer's disease but also for use against other brain proteinopathies that are exacerbated by ageing.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0368-8>.

Received: 29 September 2017; Accepted: 15 June 2018;

Published online 25 July 2018.

- Hašek, M., Chutna, J., Sládeček, M. & Lodin, Z. Immunological tolerance and tumor allografts in the brain. *Nature* **268**, 68–69 (1977).
- Louveau, A., Harris, T. H. & Kipnis, J. Revisiting the mechanisms of CNS immune privilege. *Trends Immunol.* **36**, 569–577 (2015).
- Kipnis, J. Multifaceted interactions between adaptive immunity and the central nervous system. *Science* **353**, 766–771 (2016).
- Louveau, A. et al. Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337–341 (2015).
- Aspelund, A. et al. A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules. *J. Exp. Med.* **212**, 991–999 (2015).
- Absinta, M. et al. Human and nonhuman primate meninges harbor lymphatic vessels that can be visualized noninvasively by MRI. *eLife* **6**, e29738 (2017).
- Deane, R. et al. apoE isoform-specific disruption of amyloid β peptide clearance from mouse brain. *J. Clin. Invest.* **118**, 4002–4013 (2008).
- Zhao, Z. et al. Central role for PICALM in amyloid- β blood–brain barrier transcytosis and clearance. *Nat. Neurosci.* **18**, 978–987 (2015).
- Deane, R. et al. RAGE mediates amyloid- β peptide transport across the blood–brain barrier and accumulation in brain. *Nat. Med.* **9**, 907–913 (2003).
- Mildner, A. et al. Distinct and non-redundant roles of microglia and myeloid subsets in mouse models of Alzheimer's disease. *J. Neurosci.* **31**, 11159–11171 (2011).
- Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290 (2017).
- Iliff, J. J. et al. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci. Transl. Med.* **4**, 147ra111 (2012).
- Kress, B. T. et al. Impairment of paravascular clearance pathways in the aging brain. *Ann. Neurol.* **76**, 845–861 (2014).
- Peng, W. et al. Suppression of lymphatic fluid transport in a mouse model of Alzheimer's disease. *Neurobiol. Dis.* **93**, 215–225 (2016).
- Brookmeyer, R., Abdalla, N., Kawas, C. H. & Corrada, M. M. Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. *Alzheimers Dement.* **14**, 121–129 (2018).
- Erkkinen, M. G., Kim, M. O. & Geschwind, M. D. Clinical neurology and epidemiology of the major neurodegenerative diseases. *Cold Spring Harb. Perspect. Biol.* **10**, a033118 (2018).
- Benilova, I., Karran, E. & De Strooper, B. The toxic A β oligomer and Alzheimer's disease: an emperor in need of clothes. *Nat. Neurosci.* **15**, 349–357 (2012).
- Joachim, C. L., Duffy, L. K., Morris, J. H. & Selkoe, D. J. Protein chemical and immunocytochemical studies of meningo-vascular β -amyloid protein in Alzheimer's disease and normal aging. *Brain Res.* **474**, 100–111 (1988).
- Xu, Z. et al. Deletion of aquaporin-4 in APP/PS1 mice exacerbates brain A β accumulation and memory deficits. *Mol. Neurodegener.* **10**, 58 (2015).
- Chevalier, S., Ferland, G. & Tuchweber, B. Lymphatic absorption of retinol in young, mature, and old rats: influence of dietary restriction. *FASEB J.* **10**, 1085–1090 (1996).
- Hos, D., Bachmann, B., Bock, F., Onderka, J. & Cursiefen, C. Age-related changes in murine limbal lymphatic vessels and corneal lymphangiogenesis. *Exp. Eye Res.* **87**, 427–432 (2008).
- Nagai, T., Bridenbaugh, E. A. & Gashev, A. A. Aging-associated alterations in contractility of rat mesenteric lymphatic vessels. *Microcirculation* **18**, 463–473 (2011).
- Tammela, T. et al. Photodynamic ablation of lymphatic vessels and intralymphatic cancer cells prevents metastasis. *Sci. Transl. Med.* **3**, 69ra11 (2011).
- Kilarski, W. W. et al. Optimization and regeneration kinetics of lymphatic-specific photodynamic therapy in the mouse dermis. *Angiogenesis* **17**, 347–357 (2014).
- Escobedo, N. et al. Restoration of lymphatic function rescues obesity in Prox1-haploinsufficient mice. *JCI Insight* **1**, e85096 (2016).
- Ringstad, G., Vatnehol, S. A. S. & Eide, P. K. Lymphatic MRI in idiopathic normal pressure hydrocephalus. *Brain* **140**, 2691–2705 (2017).
- Zeppenfeld, D. M. et al. Association of perivascular localization of aquaporin-4 with cognition and Alzheimer disease in aging brains. *JAMA Neurol.* **74**, 91–99 (2017).
- Rudy, J. W., Huff, N. C. & Matus-Amat, P. Understanding contextual fear conditioning: insights from a two-process model. *Neurosci. Biobehav. Rev.* **28**, 675–685 (2004).
- Owen, S. F. et al. Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature* **500**, 458–462 (2013).
- Zhang, G. et al. Hypothalamic programming of systemic ageing involving IKK- β , NF- κ B and GnRH. *Nature* **497**, 211–216 (2013).
- Liebl, J. et al. Cdk5 controls lymphatic vessel development and function by phosphorylation of Foxc2. *Nat. Commun.* **6**, 7274 (2015).
- Jeltsch, M. et al. CCBE1 enhances lymphangiogenesis via A disintegrin and metalloprotease with thrombospondin motifs-3-mediated vascular endothelial growth factor-C activation. *Circulation* **129**, 1962–1971 (2014).
- Shin, J. W. et al. Prox1 promotes lineage-specific expression of fibroblast growth factor (FGF) receptor-3 in lymphatic endothelium: a role for FGF signaling in lymphangiogenesis. *Mol. Biol. Cell* **17**, 576–584 (2006).
- Saaristo, A. et al. Lymphangiogenic gene therapy with minimal blood vascular side effects. *J. Exp. Med.* **196**, 719–730 (2002).
- Karkkainen, M. J. et al. A model for gene therapy of human hereditary lymphedema. *Proc. Natl Acad. Sci. USA* **98**, 12677–12682 (2001).
- Joukov, V. et al. A recombinant mutant vascular endothelial growth factor-C that has lost vascular endothelial growth factor receptor-2 binding, activation, and vascular permeability activities. *J. Biol. Chem.* **273**, 6599–6602 (1998).
- Han, J. et al. Vascular endothelial growth factor receptor 3 controls neural stem cell activation in mice and humans. *Cell Rep.* **10**, 1158–1172 (2015).
- Meshi, D. et al. Hippocampal neurogenesis is not required for behavioral effects of environmental enrichment. *Nat. Neurosci.* **9**, 729–731 (2006).
- Harris, J. A. et al. Many neuronal and behavioral impairments in transgenic mouse models of Alzheimer's disease are independent of caspase cleavage of the amyloid precursor protein. *J. Neurosci.* **30**, 372–381 (2010).
- Palop, J. J. et al. Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron* **55**, 697–711 (2007).
- Oakley, H. et al. Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).
- Sagare, A. P. et al. Pericyte loss influences Alzheimer-like neurodegeneration in mice. *Nat. Commun.* **4**, 2932 (2013).
- Sevigny, J. et al. Addendum: The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* **546**, 564 (2017).

Acknowledgements We thank S. Smith for editing the manuscript, J. Roy for MRI expertise, N. Al Hamadani for animal care, G. Oliver (Feinberg School of Medicine, Northwestern University, Chicago) for *Prox1*^{+/−} mice. This work was supported by grants from the National Institutes of Health/National Institute on Aging (AG034113 and AG057496), the Cure Alzheimer's Fund, Owens Family Foundation and the Thomas H. Lowder Family Foundation (awarded to J.K.), the Hobby Foundation (awarded to A.V. and S.T.A.) and American Cancer Society (IRG 81-001-26 awarded to J.M.M.). We thank all members of the Kipnis Laboratory and the BIG center for their valuable comments during numerous discussions of this work.

Reviewer information Nature thanks D. Holtzman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.D.M. designed and performed the experiments, analysed and interpreted the data and wrote the manuscript; A.L. designed and performed the experiments and participated in manuscript preparation; A.V. developed the software (Lymph4D) for MRI data processing and analysis; I.S. performed surgeries and behavioural testing; C.C. performed T2-weighted MRI, magnetic resonance angiography (MRA) and magnetic resonance venography (MRV) acquisition and data analysis; R.C.C. made hydrogels for transcranial peptide delivery; K.M.K. carried out brain T1-weighted MRI acquisition; S.O.-G. and E.F. carried out RNA-seq experiments; D.R., K.E.V., R.D.P., W.B., N.D. and R.B. assisted with experimental procedures; R.C. and S.H. carried out photoacoustic imaging; S.S.R. and J.M.M. provided resources and were involved in experimental design; M.B.L. provided human tissue samples; C.C.O. helped with RNA-seq raw data analysis, data interpretation and manuscript writing; S.T.A. participated in the development of the software (Lymph4D), provided resources and intellectual contributions; J.K. designed the experiments, provided intellectual contributions, oversaw data analysis and interpretation and wrote the manuscript.

Competing interests J.K. is an Advisor to PureTech Health/Ariya.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0368-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0368-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.D.M. and J.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mouse strains and housing. Male or female wild-type mice (C57BL/6J background) were bred in-house, purchased from the Jackson Laboratory or provided by the National Institutes of Health/National Institute on Ageing. All mice were maintained in the animal facility for habituation for at least one week before the start of the manipulation/experimentation. C57BL/6J wild-type mice were tested at 2–3, 12–14 and 20–24 months of age. Male hemizygous B6.Cg-Tg(*PDGFB-APPSwInd*)20Lms/2Mmjax (J20, JAX 006293) and B6.Cg-Tg(*APPSwFLLon,PS EN1**M146L*L286V)6799Vas/Mmjax (5xFAD, JAX 008730) were purchased from the Jackson Laboratory and bred in-house on a C57BL/6J background. J20 hemizygous mice present diffuse amyloid- β deposition in the dentate gyrus and neocortex at 5–7 months, with all transgenic mice exhibiting plaques by the age of 8–10 months⁴⁴. 5xFAD hemizygous mice overexpress the transgene constructs under neural-specific elements of the mouse thymocyte differentiation antigen 1 promoter and have accelerated accumulation of 42-residue amyloid- β peptides (amyloid- β_{42}) and deposition of amyloid and gliosis in the brain starting at two months of age, with marked amyloid plaque load without major behavioural deficits at five months⁴⁵. In-house bred male transgene carriers and non-carrier (wild-type) littermates were used at different ages that are indicated throughout the manuscript. *Prox1*^{LacZ} mice (designated *Prox1*^{+/+} mice in this manuscript) on a NMRI background (provided by G. Oliver, Northwestern University, Chicago) were also bred in-house and used in this study as a constitutive model for dysfunctional lymphatic vessels⁴⁶. Mice of all strains were housed in an environment with controlled temperature and humidity, on 12 h light:dark cycles (lights on at 7:00), and fed with regular rodent's chow and sterilized tap water ad libitum. All experiments were approved by the Institutional Animal Care and Use Committee of the University of Virginia.

Intra-cisterna magna injections. Mice were anaesthetized by intraperitoneal (i.p.) injection of a mixed solution of ketamine (100 mg kg⁻¹) and xylazine (10 mg kg⁻¹) in saline. The skin of the neck was shaved and cleaned with iodine and 70% ethanol, ophthalmic solution placed on the eyes to prevent drying and the head of the mouse was secured in a stereotaxic frame. After making a skin incision, the muscle layers were retracted and the cisterna magna exposed. Using a Hamilton syringe (coupled to a 33-gauge needle), the volume of the desired tracer solution was injected into the CSF-filled cisterna magna compartment. For brain CSF influx and lymphatic drainage experiments, 2 or 5 μ l of Alexa Fluor 594- or 647-conjugated OVA (Thermo Fisher Scientific), at 0.5 mg ml⁻¹ in artificial CSF (597316, Harvard Apparatus UK), were injected at a rate of 2.5 μ l min⁻¹. After injecting, the syringe was left in place for additional 2 min to prevent backflow of CSF. The neck skin was then sutured, after which the mice were subcutaneously injected with ketoprofen (2 mg kg⁻¹) and allowed to recover on a heat pad until fully awake. For details regarding changes in intracranial pressure associated with this injection methodology see 'Intracranial pressure measurements' and Extended Data Fig. 2.

Intracranial pressure measurements. Mice were anaesthetized by i.p. injection with ketamine and xylazine in saline and the skin was incised to expose the skull. A 0.5-mm diameter hole was drilled in the skull above the right parietal lobe. Using a stereotaxic frame, a pressure sensor catheter (model SPR100, Millar) was inserted perpendicularly into the cortex at a depth of 1 mm. To record changes in intracranial pressure (ICP), the pressure sensor was connected to the PCU-2000 pressure control unit (Millar). For measurements of ICP while performing i.c.m. injections of 2 or 5 μ l of tracer (following the same i.c.m. injection procedure as describe above), after stabilization of the signal (around a minute after insertion of the probe), average pressure was calculated over 1 min right before start injecting (pre-injection), over the last min of injection (during injection), over the last min of extra time used to prevent CSF backflow (post-injection with syringe in) and over the last 2 min of recording, specifically between minute 4 and 6 after taking out the syringe (post-injection with syringe out). For measurements in non-injected mice or in mice at different time-points (30, 60 and 120 min post-injection) after i.c.m. injection of 2, 5 or 10 μ l of tracer, ICP was recorded for 6 min after stabilization of the signal and the average pressure was calculated over the last 2 min of recording (between minute 4 and 6 of the recording). All animals were euthanized at the conclusion of the measurement.

Meningeal lymphatic vessel ablation. Selective ablation of the meningeal lymphatic vessels was achieved by i.c.m. injection and transcranial photoconversion of visudyne (verteporfin for injection, Valeant Ophthalmics), visudyne was reconstituted following the manufacturer's instructions and 5 μ l was injected i.c.m. following the procedure described in 'Intra-cisterna magna injections'. After 15 min, an incision was performed in the skin to expose the skull bone and visudyne was photoconverted by pointing a 689-nm-wavelength non-thermal red light (Coherent Opal Photoactivator, Lumenis) to five different spots above the intact skull (1 on the injection site, 1 on the superior sagittal sinus, 1 at the junction of all sinuses and 2 on the transverse sinuses). Each spot was irradiated with a light dose of 50 J per cm² at an intensity of 600 mW per cm² for a total of 83 s. Controls were injected with the same volume of visudyne (without the photoconversion step) or

sterile saline plus photoconversion (vehicle/photoconversion). The scalp skin was then sutured, after which the mice were subcutaneously injected with ketoprofen (2 mg kg⁻¹) and allowed to recover on a heat pad until fully awake.

Lymphatic vessel ligation. Surgical ligation of the lymphatics afferent to the dCLNs was performed as described previously⁴⁷. In brief, mice were anaesthetized by i.p. injection with ketamine and xylazine in saline, the skin of the neck was shaved and cleaned with iodine and 70% ethanol and ophthalmic solution placed on the eyes to prevent drying. A midline incision was made 5 mm superior to the clavicle. The sternocleidomastoid muscles were retracted and the dCLNs were exposed on each side. Ligation of the afferent lymphatic vessels on each side was performed with 10-0 synthetic, non-absorbable sutures. Control mice were subjected to a sham surgery consisting of the skin incision and retraction of the sternocleidomastoid muscle only. The skin was then sutured, after which the mice were subcutaneously injected with ketoprofen (2 mg kg⁻¹) and allowed to recover on a heat pad until fully awake.

Brain parenchymal injections. Mice were anaesthetized by i.p. injection of ketamine and xylazine in saline and the head was secured in a stereotaxic frame. An incision was made in the skin to expose the skull and a hole was drilled at 1.5 mm in the anterior–posterior axis and –1.5 mm in the medial–lateral axis relative to bregma. Then, using a Hamilton syringe (coupled to a 33-gauge needle) placed at 2.5 mm in the dorsal–ventral axis (relative to bregma), 1 μ l of Alexa Fluor 647-conjugated OVA (at 0.5 mg ml⁻¹), HiLyte Fluor 647-conjugated amyloid- β_{42} (at 0.05 μ g ml⁻¹, AnaSpec, Inc.) or BODIPY FL-conjugated low-density lipoprotein (LDL) from human plasma (at 0.1 mg ml⁻¹, Thermo Fisher Scientific) in artificial CSF were injected at a rate of 0.2 μ l min⁻¹ into the brain parenchyma. Concentrations of the injected fluorescent amyloid- β_{42} and LDL molecular tracers were chosen in order to be comparable to levels detected in the brain ISF of transgenic mice with Alzheimer's disease⁴⁷ and in plasma of C57BL/6 mice⁴⁸, respectively. After injecting, the syringe was left in place for additional 5 min to prevent backflow. The scalp skin was then sutured, after which the mice were subcutaneously injected with ketoprofen (2 mg kg⁻¹) and allowed to recover on a heat pad until further experiments.

AAV delivery. For experiments in which the effect of viral-mediated expression of mVEGF-C (NM_009506.2) on meningeal lymphatic vessels was assessed, 2 μ l of artificial CSF containing 10¹³ genome copies per ml of AAV1-CMV-mVEGF-C, or control AAV1-CMV-eGFP (AAV1, adeno-associated virus serotype 1; CMV, cytomegalovirus promoter; eGFP, enhanced green fluorescent protein; purchased from Vector BioLabs, Philadelphia), were injected directly into the cisterna magna CSF at a rate of 2 μ l min⁻¹, following the procedure described in 'Intra-cisterna magna injections'.

Transcranial recombinant VEGF-C delivery. A hydrogel of 1.4% hyaluronic acid and 3% methylcellulose alone (vehicle) or with 200 ng ml⁻¹ of encapsulated human VEGF-C (PeproTech) or VEGF-C156S (R&D Systems) was prepared as described elsewhere⁴⁹. In brief, lyophilized, sterile methylcellulose (4000 cP, Sigma-Aldrich) and sterile hyaluronic acid (1,500–1,800 kDa, Sigma-Aldrich) were sequentially dissolved in sterile 0.1 M phosphate buffered saline (PBS) at 4°C overnight. Lyophilized VEGF-C or VEGF-C156S were resuspended as particulate at 2,000 ng ml⁻¹ in 0.5% sterile methylcellulose in PBS. The particulate solution, or vehicle (0.5% methylcellulose), was mixed into the hydrogel pre-solution at 1:10, and loaded into a syringe for gelation at 37°C. The methylcellulose provided more stability, by promoting thermal gelation, and increased the hydrophobic properties of the gel⁴⁹, sustaining the release of VEGF-C or VEGF-C156S up to 7–10 days in vitro (verified using an ELISA for human VEGF-C, R&D Systems). The hydrogels were prepared on the day of the experiment and kept warm inside the individual syringes until applied onto the skull of the mouse. The mouse was anaesthetized by i.p. injection of ketamine and xylazine in saline and the head was secured in a stereotaxic frame. An incision was made in the scalp skin and the skull was thinned at the junction of all sinuses and above the transverse sinus. The shear-thinning properties of the polymers allowed the extrusion of 100 μ l of each hydrogel solution from the syringe into the thinned skull surface. The scalp skin was then sutured on top of the solidified hydrogel, after which the mice were subcutaneously injected with ketoprofen (2 mg kg⁻¹) and allowed to recover on a heat pad until fully awake. Taking the release kinetics of 7–10 days into account, hydrogels were reapplied, following the same methodology, two weeks after the first treatment.

MRI acquisitions and analysis. All MRI acquisitions were performed at the University of Virginia Molecular Imaging Core facilities in a 7T ClinScan system (Bruker) equipped with a 30-mm diameter cylindrical RF Coil. Detailed descriptions of MRI data acquisition, processing and analysis (including mathematical models and equations) can be found in the Supplementary information.

Photoacoustic imaging. Adult mice were maintained under anaesthesia with 1.5% isoflurane and at a constant body temperature with the aid of a heat pad. A surgical incision was made in the scalp and the fascia was removed to expose the skull. One day before the imaging, the skull over the region of interest was thinned to the desired thickness (~100 μ m). Mice were then imaged by multi-parametric

photoacoustic microscopy, which is capable of simultaneously imaging oxygen saturation of hemoglobin (SO_2) and blood flow speed as described previously⁵⁰. Using the oxy-haemoglobin and deoxy-haemoglobin values, recorded using two nanosecond-pulsed lasers (532 and 559 nm), it is possible to compute the final SO_2 . Correlation analysis of adjacent A-line signals allows the quantification of blood flow speed within individual vessels. By segmenting major vessels within the region of interest, average values of the blood flow speed and SO_2 were extracted for quantitative analysis.

Open field test. The open field test was performed following a published protocol⁵¹ with minor modifications. Mice were carried to the behaviour room to habituate at least 30 min before starting the test. Mice were then placed into the open field arena (made of opaque white plastic material, 35 cm \times 35 cm) by a blinded experimenter and allowed to explore the arena for 15 min. Total distance (cm) and percentage of time spent in the centre (22 cm \times 22 cm) were quantified using video tracking software (TopScan, CleverSys, Inc.).

NLR test. The novel location recognition test was performed following a published protocol⁵² with modifications. The experimental apparatus used in this study was the same square box made of opaque white plastic (35 cm \times 35 cm) used in the open field test. The mice were first habituated to the apparatus for 15 min. Two different plastic objects (one red and the other blue, and with different shapes) were then positioned in the arena, in two corners next to each other and 5 cm away from each adjacent arena wall. Mice were then placed in the arena (by a blinded experimenter), facing the wall furthest away from the objects and allowed to explore the arena and objects for 10 min. After 24 h, the mice were placed in the same box with the same two objects, but one of them had switched location and was placed in a new quadrant, obliquely to the familiar object (novel location test). The time spent exploring the objects in the familiar and novel locations was also measured for 10 min. Exploration of an object was assumed when the mouse approached an object and touched it with its vibrissae, snout or forepaws and was measured using a video tracking software (TopScan, CleverSys, Inc.). The object location preference (percentage of time with object) was calculated as the exploration time of the objects in the familiar or in the novel location/total exploration time.

CFC test. This behavioural test was performed following a published protocol⁵³ with modifications. In this associative learning task, mice were presented with a neutral conditioned cue stimulus that is paired with an aversive unconditioned stimulus in a particular context. The mice learned that the chamber context and the cue stimulus predicted the aversive stimulus and this elicited a specific behavioural response, namely freezing. Mice were brought into the testing room to acclimatize for at least 30 min before testing. For the test, we used two Habitest chambers (Coulbourn Instruments) with stainless grid floors attached to a shock generator for foot shock delivery and dimly illuminated with a white-fluorescent light bulb. The chambers were cleaned and made odour-free before starting the experiment and between each session (or each mouse). The fear conditioning test was conducted over two days. On day 1, mice were placed in the conditioning chamber and allowed to habituate for 3 min. Then, mice received three pairs of cue-aversive stimuli, consisting of tone (18 s, 5 kHz, 75 dB)–shock (2 s, 0.5 mA) pairings, separated by an interval of 40 s (total of 3 min). Mice were returned to their home cage 30 s after the last shock presentation. On day 2, mice were tested and scored for conditioned fear to the training context for 3 min (context test), but with no presentation of the cue stimulus. After 2 h, mice were presented to a novel context, in which the light intensity was slightly increased, the grid and walls of the chamber were covered by plastic inserts with different texture and colours and the inside of the chamber was scented with a paper towel dabbed with vanilla extract placed under the floor grid. In this last session, mice were placed in the conditioning chamber and allowed to habituate for 3 min, after which they received a continuous cue stimulus (tone) for an additional 3 min (cued test). Mice behaviour was recorded by a digital video camera mounted above the conditioning chamber and freezing was manually scored by a blinded experimenter using the Etholog v.2.2 software. Parameters analysed included the percentage of time freezing during the 3 min of the context test and the last 3 min of the cued test.

MWM test. The MWM test was performed as described previously⁵³, but with modifications. Mice were transported to the behaviour room to habituate at least 30 min before starting the test. The MWM test consisted of four days of acquisition, one day of probe trial and two days of reversal. In the acquisition, mice performed four trials per day, for four consecutive days, to find a hidden 10-cm diameter platform located 1 cm below the water surface in a pool that was 1 m in diameter. Tap water was made opaque with nontoxic tempera white paint and the water temperature was kept at $23 \pm 1^\circ\text{C}$. A dim light source was placed within the testing room and only distal visual cues were available above each quadrant of the swimming pool to aid in the spatial navigation and location of the submerged platform. The latency to platform, that is, the time required by the mouse to find and climb onto the platform, was recorded for up to 60 s. Each mouse was allowed to remain on the platform for 20 s and was then moved from the maze to its home cage. If the mouse did not find the platform within 60 s, it was manually placed

on the platform and returned to its home cage after 20 s. The inter-trial interval for each mouse was at least 5 min. On day 5, the platform was removed from the pool, and each mouse was tested in a probe trial for 60 s. On days 1 and 2 of the reversal, without changing the position of the visual cues, the platform was placed in the quadrant opposite to the original acquisition quadrant and the mouse was retrained for four trials per day. All MWM testing was performed between 13:00 and 18:00, during the lights-on phase, by a blinded experimenter. During the acquisition, probe and reversal, data were recorded using the EthoVision automated tracking system (Noldus Information Technology). The mean latency (in seconds) of the four trials was calculated for each day of test trials. The percentage of time in the platform quadrant was calculated for the probe trial. Additionally, using a modified version of previous published methods^{54,55}, the full tracked path taken by each mouse in every trial of the acquisition and reversal days was used to classify the type of navigation strategy as either egocentric or allocentric by a blinded experimenter. The mean percentage of allocentric navigation of four trials was calculated for each day.

CSF and tissue collection and processing. Mice were given a lethal dose of anaesthetics by i.p. injection of euthasol (10% v/v in saline). When needed, CSF was collected from the cisterna magna using a 0.5-mm diameter borosilicate glass pipette with internal filament and immediately stored at -80°C . Mice were then transcardially perfused with ice-cold PBS with heparin (10 U ml^{-1}). Deep cervical lymph nodes were dissected and drop-fixed in 4% paraformaldehyde (PFA) for 12 h at 4°C . After stripping the skin and muscle from the bone, the head was collected and drop-fixed in 4% PFA. After removal of the mandibles and the skull rostral to maxillae, the top of the skull (skullcap) was removed with surgical curved scissors by cutting clockwise, beginning and ending inferior to the right post-tympanic hook and kept in PBS and 0.02% azide at 4°C until further use. The brains were kept in 4% PFA for an additional 24 h (48 h in total). Fixed brain and dCLNs were then washed with PBS, cryoprotected with 30% sucrose and frozen in Tissue-Plus OCT compound (Thermo Fisher Scientific). Fixed and frozen brains were sliced ($100\text{-}\mu\text{m}$ thick sections) with a cryostat (Leica) and kept in PBS and 0.02% azide at 4°C . Frozen lymph nodes were sliced ($30\text{-}\mu\text{m}$ thick sections) in a cryostat, collected onto gelatin-coated Superfrost Plus slides (Thermo Fisher Scientific) and stored at -20°C . Alternatively, after euthanizing and perfusing the mouse, the skullcap was removed from the head of the mouse and drop-fixed in 4% PFA for 12 h, and the brains were immediately collected into OCT compound, snap-frozen in dry ice and stored at -80°C . Fresh-frozen brains were then sliced ($30\text{-}\mu\text{m}$ thick sections) in the cryostat and sections were directly collected onto Superfrost Plus slides and kept at -20°C until further use. Fixed meninges (dura mater and arachnoid) were carefully dissected from the skullcaps with Dumont #5 forceps (Fine Science Tools) and kept in PBS and 0.02% azide at 4°C until further use.

Amyloid- β measurement in CSF. To measure the concentration of amyloid- β peptides (ranging in size from amyloid- β_{37} to amyloid- β_{42}) in the CSF of J20 mice, an in-house direct ELISA assay was used. In brief, Nunc MaxiSorp flat-bottom 96-well plates (ThermoFisher Scientific) were coated with $2\text{ }\mu\text{l}$ of CSF diluted in $98\text{ }\mu\text{l}$ of a $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ buffer (pH 8.0) solution (1:50 dilution factor), for 2 h at 37°C . After washing with PBS and 0.05% Tween-20 (Sigma-Aldrich), a blocking step with PBS and 1% skim milk was performed for 1 h at room temperature. Then, consecutive incubations for 1 h at room temperature were performed: first, with rabbit anti-amyloid- β_{37-42} (Cell Signaling, clone D54D2, 1:500); second, with biotinylated goat anti-rabbit (Vector Laboratories, BA-1000, 1:500); and third, with streptavidin-horseradish peroxidase (1:2,500, Sigma-Aldrich). Each incubation step was separated by two washes with PBS containing 0.05% Tween 20 and followed by another two washes with PBS alone. Finally, a citrate-phosphate buffer (pH 4.3) solution containing 0.1% of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS, Sigma-Aldrich) was added to each well and absorbance was read at 405 nm. The standard curve used to extrapolate the concentration of amyloid- β in the CSF was obtained using known concentrations of human amyloid- β_{42} (AnaSpec, Inc.) that ranged from 0.1 to 100 ng ml^{-1} (considering the linearity of the assay). Data processing was done with Excel and statistical analysis was performed using Prism 7.0a (GraphPad Software).

Human samples. Autopsy specimens of human brain and dura from patients without ($n=8$) or with ($n=9$) Alzheimer's disease were obtained from the Department of Pathology at the University of Virginia (UVA). All samples were from consenting patients that gave no restriction to the use of their body for research and teaching (through an UVA's Institutional Review Board for Health Sciences Research). Diagnosis criteria and pathological score were performed following the National Institute on Ageing–Alzheimer's Association guidelines⁵⁶, based on the ABC (Amyloid, Braak, CERAD) score, for seven of the cases with Alzheimer's disease; old guidelines were used to diagnose and score two of the cases with Alzheimer's disease (Extended Data Table 1). All obtained samples were fixed in a 20% formalin solution and kept in paraffin blocks until further sectioning. Prior to immunohistochemical staining, slides containing $10\text{-}\mu\text{m}$ thick sections were heated to 70°C for 30 min and deparaffinized by washing sections with xylene, 1:1 xylene:100%

ethanol (v/v), and 100, 95, 70 and 50% ethanol in water. Finally, tissue sections were rehydrated by rinsing with cold tap water.

Immunohistochemistry, imaging and quantifications. Mouse fresh-frozen brain sections were fixed with 4% PFA for 30 min, rinsed in dH₂O and subjected to a heat-induced antigen retrieval step with 10 mM citrate buffer for 20 min. After deparaffinization, sections of human brain or dura were subjected to the same antigen retrieval step for 20 min. The steps described next were applied for mouse fresh-frozen and fixed free-floating brain sections, lymph node sections on slide, meningeal whole-mounts and human fixed tissue. For immunofluorescence staining, tissue was rinsed in PBS and washed with PBS and 0.5% Triton X-100 for 10 min, followed by incubation in PBS and 0.5% Triton X-100 containing 0.5% of normal serum (either goat or chicken) and 0.5% bovine serum albumin (BSA) for 1 h at room temperature. This blocking step was followed by incubation with appropriate dilutions of primary antibodies: anti-LYVE-1-eFluor 660 or anti-LYVE-1-Alexa Fluor 488 (eBioscience, clone ALY7, 1:200), anti-CD31 (Millipore Sigma, MAB1398Z, clone 2H8, 1:200), anti-IBA1 (Abcam, ab5076, 1:300), anti-GFAP (Millipore Sigma, ab5541, 1:300), anti-AQP4 (Millipore Sigma, A5971, 1:200), anti-Ki-67 (Abcam, ab15580, 1:100), anti-human amyloid- β_{16} (BioLegend, clone 6E10, 1:200), anti-amyloid- β_{37-42} (Cell Signaling, clone D54D2, 1:300) and anti-GFP (Abcam, ab6556, 1:300) in PBS and 0.5% Triton X-100 containing 0.5% of normal serum and 0.5% BSA overnight at 4°C. Meningeal whole-mounts or tissue sections were then washed three times for 5 min at room temperature in PBS and 0.5% Triton X-100 followed by incubation with the appropriate chicken, goat or donkey Alexa Fluor 488, 546, 594 or 647 anti-rat, -goat, -rabbit, -mouse or -Armenian hamster IgG antibodies (Thermo Fisher Scientific, 1:500) for 1 or 2 h at room temperature in PBS and 0.5% Triton X-100. After incubating for 10 min with 1:2,000 DAPI in PBS, the tissue was washed three times for 5 min with PBS at room temperature and mounted with Aqua-Mount (Lerner) and glass coverslips. Preparations were stored at 4°C for no more than one week until images were acquired either using a wide-field microscope (Leica) or a confocal microscope (FV1200 Laser Scanning Confocal Microscope, Olympus). Quantitative analysis using the acquired images was performed using Fiji software. For the assessment of brain fluorescent tracer influx or efflux or AQP4 coverage, 10 representative brain sections were imaged using the wide-field microscope and the mean area fraction was calculated using Microsoft Excel. For lymph nodes, the area fraction of drained fluorescent tracer or lymphatic vessels was assessed in alternate sections (representing a total of 10–15 sections per sample) using a confocal microscope and the mean was calculated for each sample. Area of coverage by CD31⁺ blood vessels and AQP4⁺ astrocyte endfeet in the brain cortex was achieved by calculating the mean value of 10 representative fields (5 images in each cerebral hemisphere) per sample that was acquired using a confocal microscope. For lymphatic vessel diameter, images of the same region of the superior sagittal sinus or of the transverse sinus were acquired using a confocal microscope and the mean of 100 individual lymphatic vessel diameter measurements (50 measurements in each lymphatic vessel lining the sinus using Fiji) was calculated for each sample by a blinded experimenter (due to different criteria used by distinct experimenters, this quantification method is often associated with a variability of $\pm 15\%$ in absolute diameter values). For assessment of meningeal lymphatic vessel coverage and complexity, images of meningeal whole-mounts were acquired using a confocal microscope and Fiji was used for quantifications. When applicable, the same images were used to assess the percentage of field coverage by LYVE-1⁺CD31⁺ vessels. To quantify the number of proliferating Ki-67⁺ cells in the hippocampal dentate gyrus, images of the entire dentate gyrus of three representative brain sections per sample were obtained using a confocal microscope. Fiji was used to assess the number of Ki-67⁺ cells per mm² of DAPI cells that composed the granular zone, which were then used to calculate the average density of cells per sample. For assessment of amyloid burden in the dorsal hippocampus, tile scans of the entire dorsal hippocampus from 10 coronal brain sections ($\sim 180\mu\text{m}$ apart from each other) were obtained using a confocal microscope. Fiji was used to quantify amyloid plaque size, number and total coverage.

Flow cytometry. Mice were injected i.p. with euthasol solution and were then transcardially perfused with ice-cold PBS with heparin. Individual meninges were immediately dissected from the skullcap of the mouse and digested for 15 min at 37°C with 1.4 U ml⁻¹ of collagenase VIII (Sigma Aldrich) and 35 U ml⁻¹ of DNase I (Sigma Aldrich) in complete media consisting of DMEM (Gibco) with 2% FBS (Atlas Biologicals), 1% L-glutamine (Gibco), 1% penicillin-streptomycin (Gibco), 1% sodium pyruvate (Gibco), 1% non-essential amino-acids (Gibco) and 1.5% HEPES (Gibco). The cell pellets were washed, resuspended in ice-cold fluorescence-activated cell sorting (FACS) buffer (pH 7.4; 0.1 M PBS; 1 mM EDTA and 1% BSA) and stained for extracellular markers with the following antibodies: rat anti-CD90.2-FITC (553013; BD Biosciences), rat anti-CD11b-FITC (557396; BD Biosciences), rat monoclonal anti-CD19-PE (12-0193-82; eBioscience), rat anti-CD45-PerCP-Cy5.5 (550994; BD Biosciences), rat anti-Ly6C-PerCP-Cy5.5 (56

0525; BD Biosciences), mouse anti-NK1.1-PE-Cy7 (552878; BD Biosciences), rat anti-Ly6G-PE-Cy7 (560601; BD Biosciences), rat anti-CD4-APC (553051; BD Biosciences), rat anti-CD45-Alexa Fluor 700 (560510; BD Biosciences), hamster anti-TCRb-BV711 (563135; BD Biosciences), rat anti-CD8-Pacific blue (558106; BD Biosciences) and rat anti-Siglec-F-BV421 (562681; BD Biosciences). Cell viability was determined by using the Zombie Aqua Fixable Viability Kit following the manufacturer's instructions (BioLegend). After an incubation period of 30 min at 4°C, cells were washed and fixed in 1% PFA in PBS. Fluorescence data were collected with a Gallios Flow Cytometer (Beckman Coulter, Inc.) then analysed using FlowJo software (Tree Star, Inc.). In brief, singlets were gated using the height, area and the pulse width of the forward and side scatter and then viable cells were selected as AQUA⁺. Cells were then gated for the appropriate cell-type markers. An aliquot of unstained cells of each sample was counted using Cellometer Auto2000 (Nexcelor) to provide accurate counts for each population. Data processing was done with Excel and statistical analysis was performed using Prism 7.0a (GraphPad Software, Inc.).

Sorting of meningeal LECs. To obtain a suspension of meningeal LECs from the meninges of young-adult (2–3 months) and old (20–24 months) mice using FACS, mice were euthanized by i.p. injection of euthasol and transcardially perfused with ice-cold PBS with heparin. Skullcaps were quickly collected and meninges (dura mater and arachnoid) were dissected using Dumont #5 forceps in complete medium composed of DMEM (Gibco) with 2% FBS (Atlas Biologicals), 1% L-glutamine (Gibco), 1% penicillin-streptomycin (Gibco), 1% sodium pyruvate (Gibco), 1% non-essential amino-acids (Gibco) and 1.5% HEPES (Gibco). Individual meninges were then incubated with 1 ml of complete medium with 1.4 U ml⁻¹ of collagenase VIII (Sigma-Aldrich) and 35 U ml⁻¹ of DNase I (Sigma-Aldrich) for 15 min at 37°C. Individual samples consisted of cell suspensions pooled from 10 meninges that were obtained after filtration through a 70- μm nylon-mesh cell strainer. Cell suspensions were then pelleted, resuspended in ice-cold FACS buffer containing DAPI (1:1,000, Thermo Fisher Scientific), anti-CD45-BB515 (1:200, clone 30-F11, BD Biosciences), anti-CD31-Alexa Fluor 647 (1:200, clone 390, BD Biosciences) and anti-podoplanin-PE (1:200, clone 8.1.1, eBioscience) and incubated for 15 min at 4°C. Cells were then washed and resuspended in ice-cold FACS buffer. In brief, singlets were gated using the pulse width of the side scatter and forward scatter. Cells negative for DAPI were selected as viable cells. The LECs were then gated as CD45⁺CD31⁺podoplanin⁺ (see Extended Data Fig. 6 for representative dot plots) and sorted into a 96-well plate containing 100 μl of lysis buffer (Arcturus PicoPure RNA Isolation Kit, Thermo Fisher Scientific) using the Influx Cell Sorter (BD Biosciences) that is available at the University of Virginia Flow Cytometry Core Facility.

RNA extraction and sequencing. For total RNA extraction from the whole hippocampus, the tissue was macrodissected from the brain in ice-cold PBS, immersed in the appropriate volume of extraction buffer from the RNA isolation kit, immediately snap-frozen in dry ice and stored at -80°C until further use. After defrosting on ice, samples were mechanically dissociated in extraction buffer and RNA was isolated using the kit components according to the manufacturer's instructions (RNeasy mini kit, 74106, QIAGEN). The Illumina TruSeq Stranded Total RNA Library Prep Kit was used for cDNA library preparation from total RNA samples. Sample quality control was performed using an Agilent 4200 TapeStation Instrument, using the Agilent D1000 kit, and using the Qubit Fluorometer (Thermo Fisher Scientific). For RNA-seq, libraries were loaded on to a NextSeq 500 (Illumina) using an Illumina NextSeq High Output (150 cycle) cartridge (FC-404-2002).

Total RNA was extracted from LECs (previously sorted by FACS) using the Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific), following the manufacturer's instructions. All RNA sample processing (including linear RNA amplification and cDNA library generation) and RNA-seq was performed by HudsonAlpha Genomic Services Laboratory.

The raw sequencing reads (FASTQ files) were first chastity filtered, which removes any clusters that have a higher than expected intensity of the called base compared to other bases. The quality of the reads was then evaluated using FastQC⁵⁷, and after passing quality control, the expression of the transcripts was quantified against the UCSC mm10 genome⁵⁸ using Salmon⁵⁹. These transcript abundances were then imported into R and summarized with tximport⁶⁰, and then DESeq2⁶¹ was used to normalize the raw counts, perform exploratory analysis (for example, principal component analysis), and to perform differential expression analysis. Before differential expression analysis of the meningeal LECs from the adult versus old mice dataset, surrogate variable analysis⁶² (SVA) was used to identify and adjust for latent sources of unwanted variation as implemented in the SVA package⁶³. The *P* values from the differential expression analysis were corrected for multiple hypothesis testing with the Benjamini-Hochberg false-discovery rate procedure (adjusted *P* value). Functional enrichment of differential expressed genes, using gene sets from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), was determined with Fisher's exact test as implemented in the

clusterProfiler⁶⁴ Bioconductor package. Heat maps of the differential expressed genes and enriched gene sets were generated with the R package pheatmap⁶⁵. Normalized counts of selected transcripts were used to calculate the fold change relative to respective controls.

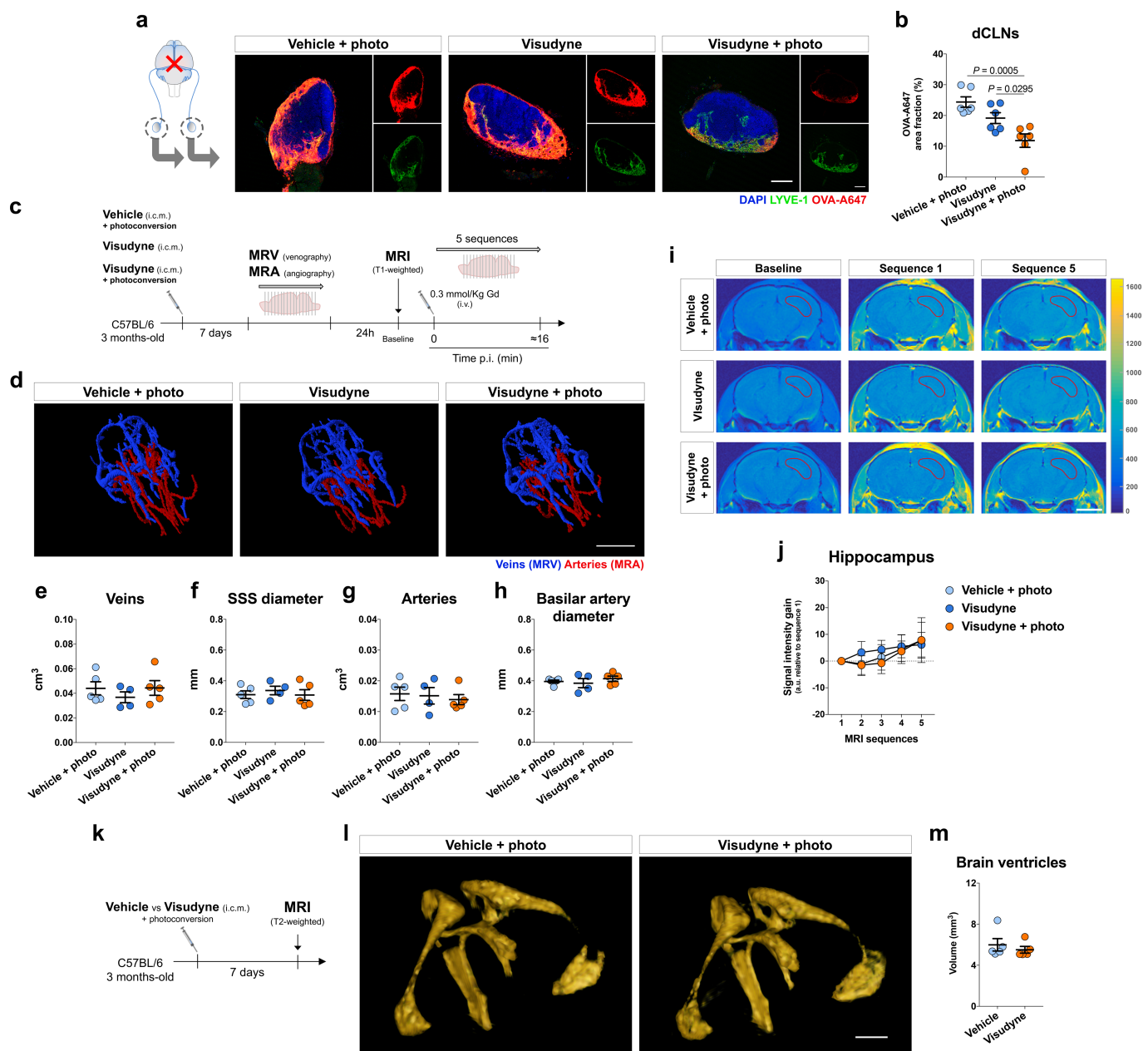
Statistical analysis and reproducibility. Sample sizes were chosen on the basis of standard power calculations (with $\alpha = 0.05$ and power of 0.8) performed for similar experiments that were previously published^{51–53}. In general, statistical methods were not used to re-calculate or predetermine sample sizes. The Kolmogorov–Smirnov test was used to assess normal distribution of the data. Variance was similar within comparable experimental groups. Animals from different cages, but within the same experimental group, were selected to assure randomization. Experimenters were blinded to the identity of experimental groups from the time of euthanasia until the end of data collection and analysis for at least one of the independent experiments. Statistical tests for each figure were justified to be appropriate. One-way ANOVA, with Bonferroni's post hoc test or Holm–Sidak's post hoc test, was used to compare three independent groups. Two-group comparisons were made using two-tailed unpaired Mann–Whitney *U*-tests. For comparisons of multiple factors (for example, age versus treatment), two-way ANOVA with Bonferroni's post hoc test was used. Repeated-measures two-way ANOVA with Bonferroni's post hoc test was used for day versus treatment comparisons with repeated observations. Statistical analysis (data are always presented as mean \pm s.e.m.) was performed using Prism 7.0a (GraphPad Software, Inc.).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Lymph4D software code is available online under GNU General Public license v.3.0 at <https://github.com/avaccari/Lymph4D>. Custom code used during the current study are also available from the corresponding authors upon reasonable request.

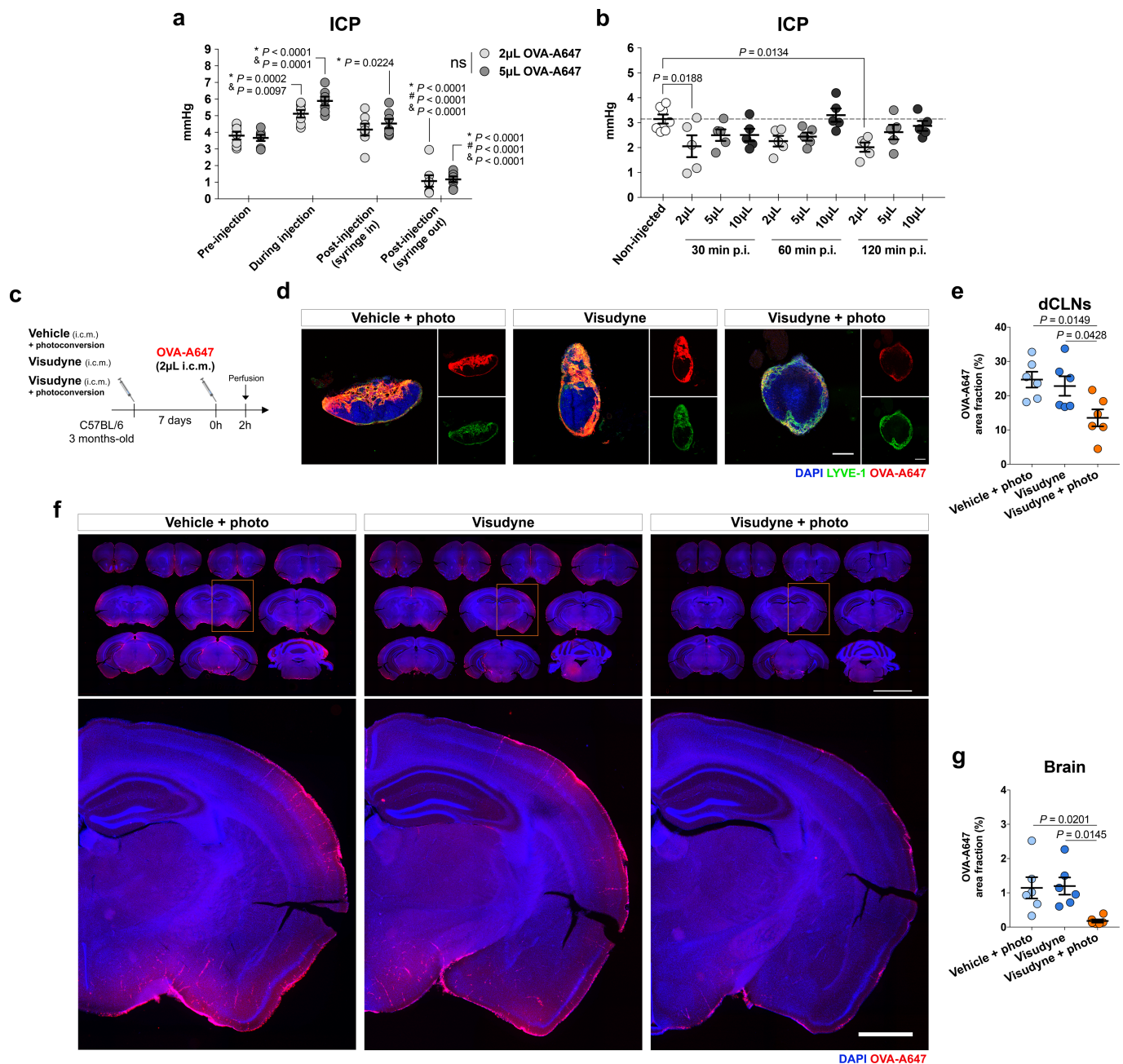
Data availability. Source Data for quantifications mentioned either in the text or shown in graphs plotted in Figs. 1–3 and Extended Data Figs. 1–9 are available in the online version of the paper. RNA-seq datasets have been deposited online in the Gene Expression Omnibus (GEO) under accession numbers GSE104181, GSE104182 and GSE113351. The datasets generated and/or analysed during the current study are also available from the corresponding authors upon reasonable request.

44. Mucke, L. et al. High-level neuronal expression of A β _{1–42} in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).
45. Richard, B. C. et al. Gene dosage dependent aggravation of the neurological phenotype in the 5XFAD mouse model of Alzheimer's disease. *J. Alzheimers Dis.* **45**, 1223–1236 (2015).
46. Harvey, N. L. et al. Lymphatic vascular defects promoted by *Prox1* haploinsufficiency cause adult-onset obesity. *Nat. Genet.* **37**, 1072–1081 (2005).
47. Cirrito, J. R. et al. In vivo assessment of brain interstitial fluid with microdialysis reveals plaque-associated changes in amyloid- β metabolism and half-life. *J. Neurosci.* **23**, 8844–8853 (2003).
48. Hinder, L. M., Vincent, A. M., Hayes, J. M., McLean, L. L. & Feldman, E. L. Apolipoprotein E knockout as the basis for mouse models of dyslipidemia-induced neuropathy. *Exp. Neurol.* **239**, 102–110 (2013).
49. Caicco, M. J. et al. A hydrogel composite system for sustained epi-cortical delivery of cyclosporin A to the brain for treatment of stroke. *J. Control. Release* **166**, 197–202 (2013).
50. Ning, B. et al. Ultrasound-aided multi-parametric photoacoustic microscopy of the mouse brain. *Sci. Rep.* **5**, 18775 (2015).
51. Derecki, N. C. et al. Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* **484**, 105–109 (2012).
52. Louveau, A. et al. Impaired spatial memory in mice lacking CD3 ζ is associated with altered NMDA and AMPA receptors signaling independent of T-cell deficiency. *J. Neurosci.* **33**, 18672–18685 (2013).
53. Derecki, N. C. et al. Regulation of learning and memory by meningeal immunity: a key role for IL-4. *J. Exp. Med.* **207**, 1067–1080 (2010).
54. Janus, C. Search strategies used by APP transgenic mice during navigation in the Morris water maze. *Learn. Mem.* **11**, 337–346 (2004).
55. Garthe, A. & Kempermann, G. An old test for new neurons: refining the Morris water maze to study the functional relevance of adult hippocampal neurogenesis. *Front. Neurosci.* **7**, 63 (2013).
56. Montine, T. J. et al. National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* **123**, 1–11 (2012).
57. Andrews, S. *FastQC: a Quality Control Tool for High Throughput Sequence Data*. Babraham Bioinformatics version 0.11.7. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
58. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
59. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
60. Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, e161 (2007).
63. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
64. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
65. Kolde, R. *Pretty Heatmaps*. R package version 1.0.8 <https://github.com/raivokolde/pheatmap> (2015).



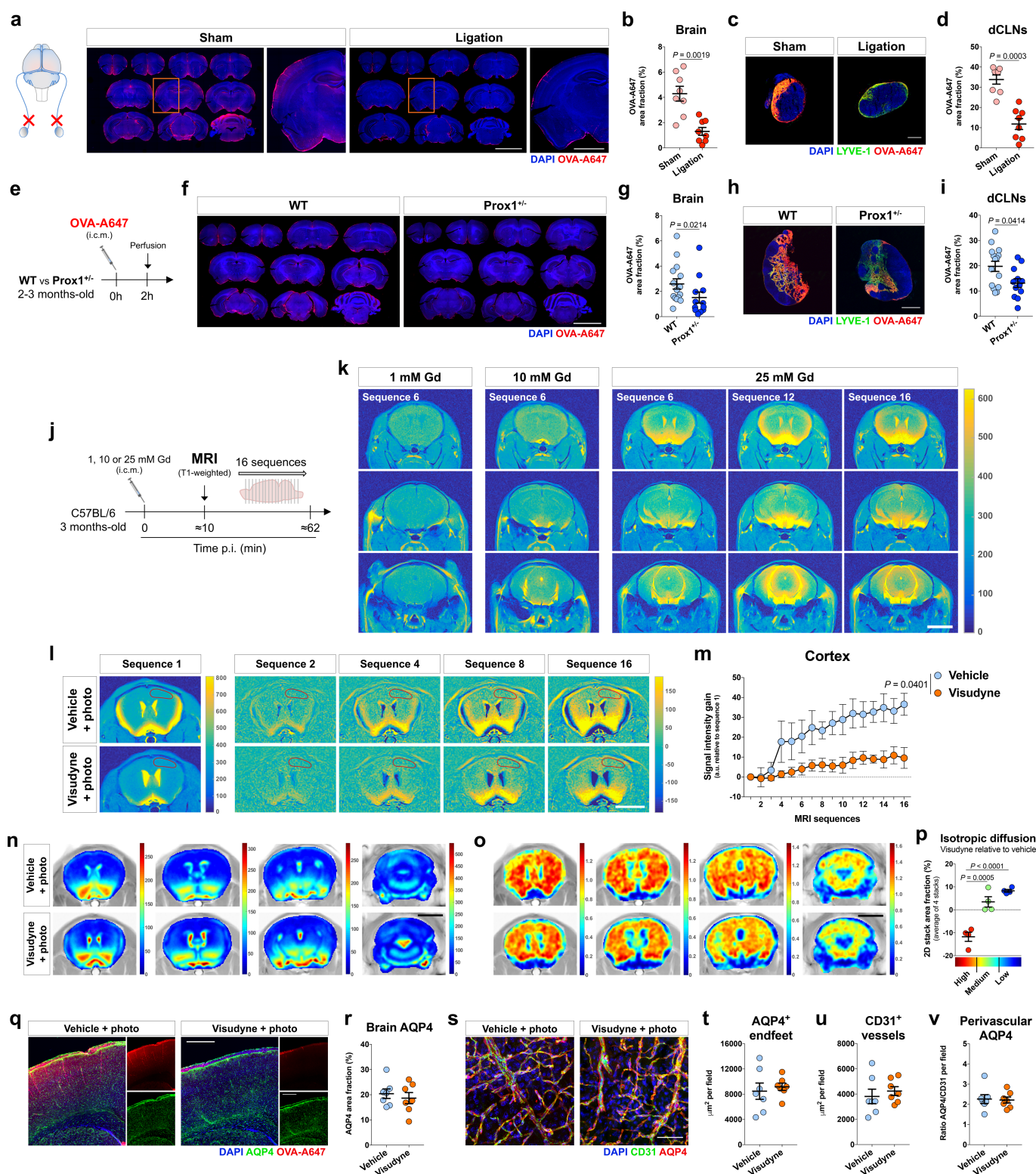
Extended Data Fig. 1 | Ablation of meningeal lymphatic vessels leads to decreased CSF macromolecule drainage affecting meningeal/brain blood vasculature or brain ventricular volume. **a**, Seven days after meningeal lymphatic ablation, a volume of 5 μ l of fluorescent OVA-A647 was injected i.c.m. into the CSF, and drainage of tracer into the dCLNs was assessed 2 h later. Representative images of OVA-A647 (red) drained into the dCLNs stained for LYVE-1 (green) and with DAPI (blue). Scale bar, 200 μ m. **b**, Quantification of OVA-A647 area fraction (%) in the dCLNs showed significantly less amount of tracer in the visudyne with photoconversion group than in control groups. Data are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test. **a, b**, Data are representative of two independent experiments; significant differences between vehicle with photoconversion and visudyne with photoconversion groups were observed in a total of five independent experiments. **c**, Seven days after meningeal lymphatic ablation, mice from the three groups were subjected to magnetic resonance venography (MRV) or angiography (MRA) and 24 h later to T2-weighted MRI to assess blood-brain barrier integrity after i.v. injection of the contrast agent Gd at a dose of 0.3 mmol kg⁻¹. **d**, Representative 3D reconstructions of intracranial veins and arteries of mice from each group. Scale bar, 5 mm.

e–h, No significant changes between groups were observed for venous vessel volume (**e**), superior sagittal sinus (SSS) diameter (**f**), arterial vessel volume (**g**) and basilar artery diameter (**h**). Data are mean \pm s.e.m., $n = 5$ in vehicle with photoconversion and in visudyne with photoconversion, $n = 4$ in visudyne; one-way ANOVA with Bonferroni's post hoc test. **i**, Using the Lymph4D software, it was possible to measure changes in signal intensity gain in MRI sequences 1–5 (relative to baseline) in the hippocampus of mice from each group. Scale bar, 3 mm. **j**, Quantification of the signal intensity gain (relative to baseline) in the hippocampus over 5 MRI acquisition sequences showed no differences between groups. Data are mean \pm s.e.m., $n = 5$ in vehicle with photoconversion and in visudyne with photoconversion, $n = 4$ in visudyne; repeated-measures two-way ANOVA with Bonferroni's post hoc test. **k**, Mice were subjected to T2-weighted MRI to assess volume changes in brain ventricles seven days after injection of vehicle or visudyne and photoconversion. **l**, Representative images of 3D reconstruction of brain ventricles of mice from the two groups. Scale bar, 1 mm. **m**, No differences were detected in the volume of the brain ventricles after meningeal lymphatic ablation. Data are mean \pm s.e.m., $n = 5$ per group, two-tailed Mann-Whitney *U*-test.



Extended Data Fig. 2 | ICP measurements and assessment of CSF drainage and brain influx. **a**, ICP was measured in four different steps of i.c.m. injection of 2 µL or 5 µL of tracer solution: pre-injection, during injection, post-injection (with syringe inside the cisterna magna) and post-injection (with syringe out of the cisterna magna). A significant increase in ICP for each volume was observed during injection when compared to pre-injection and post-injection (syringe in). Significantly higher ICP values post-injection (syringe in) were observed when compared to ICP values pre-injection. A significant decrease in ICP for each volume was observed post-injection (syringe out) when compared to all other steps of i.c.m. injection. No significant differences in ICP values were observed between groups injected with 2 µL or 5 µL of tracer for any of the analysed steps of the i.c.m. injection method. Data are mean \pm s.e.m., $n = 7$ per group; repeated-measures two-way ANOVA with Bonferroni's post hoc test, *versus pre-injection, #versus during injection, &versus post-injection (syringe in); data were pooled from two independent experiments. **b**, ICP was measured 30, 60 and 120 min post-injection (p.i.) of 2, 5 or 10 µL of tracer solution into the CSF and compared to ICP values in non-injected mice. Significant differences were observed between ICP

values of non-injected mice and mice injected with 2 µL of tracer at 30 min and 120 min post-injection. Data are mean \pm s.e.m., $n = 5$ per group, one-way ANOVA with Bonferroni's post hoc test. **c**, Seven days after meningeal lymphatic ablation, a volume of 2 µL of fluorescent OVA-A647 was injected into the CSF and drainage of tracer into the dCLNs was assessed 2 h later. **d**, Representative images of OVA-A647 (red) drained into the dCLNs, stained for LYVE-1 (green) and with DAPI (blue). Scale bars, 200 µm. **e**, Quantification of OVA-A647 area fraction (%) in the dCLNs showed significantly less amount of tracer in the visudyne with photoconversion group than in control groups. **f**, Representative brain sections stained with DAPI (blue) showing OVA-A647 (red) influx into the brain parenchyma of mice from visudyne with photoconversion and control groups. Scale bar, 5 mm and 1 mm (inset). **g**, Quantification of OVA-A647 area fraction (%) in brain sections showing a significant decrease in the visudyne with photoconversion group when compared to control groups. Data in **e** and **g** are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test; **c–g**, Data are representative of two independent experiments.

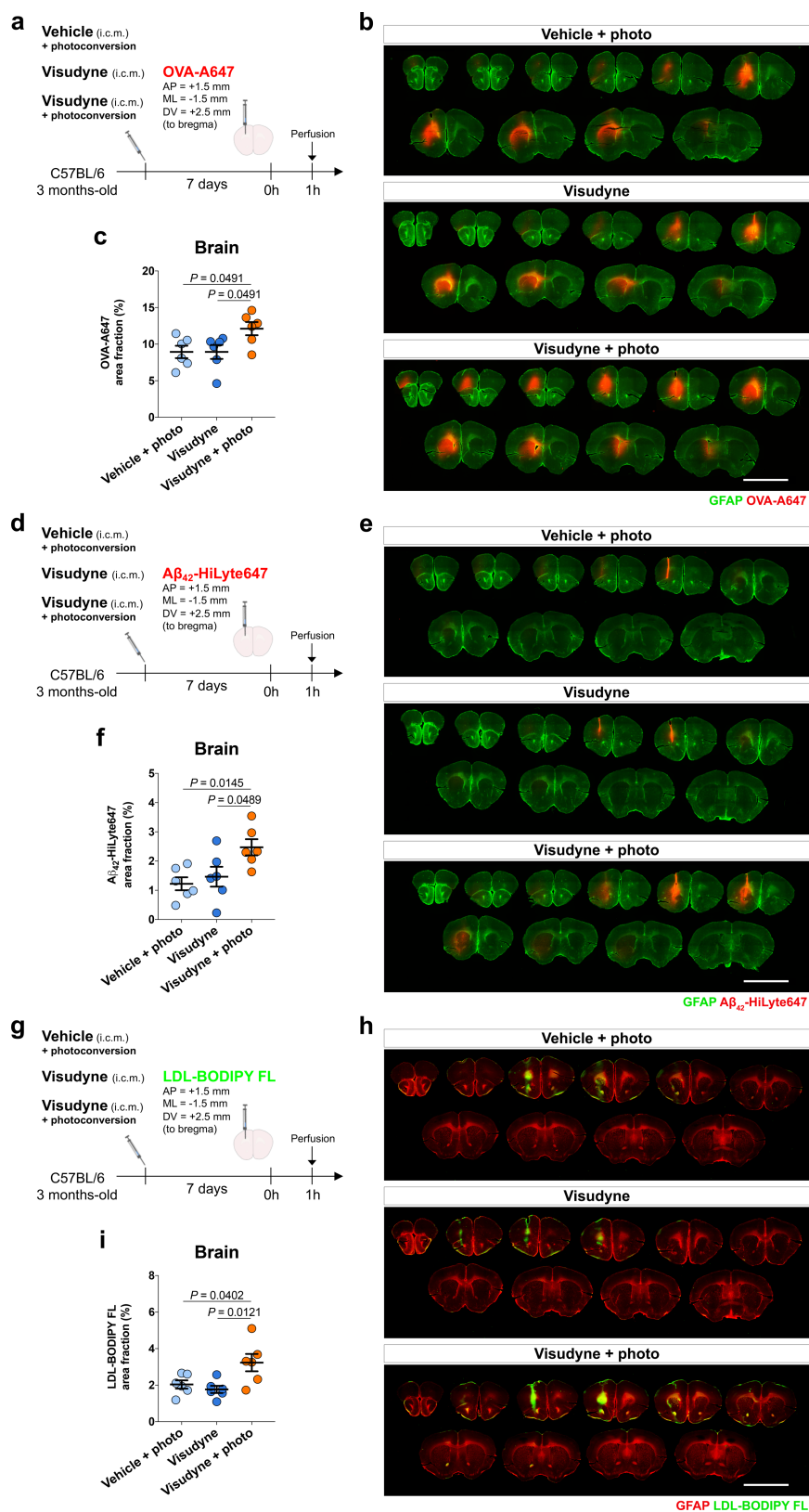


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Impaired brain perfusion by CSF**macromolecules is observed in ligated lymphatic vessels and in *Prox1*^{+/-} mice and does not correlate with AQP4 levels.**

a, Adult mice were subjected to surgical ligation of the lymphatic vessels afferent to the dCLNs. One week after the procedure, 5 μ l of OVA-A647 was injected into the CSF (i.c.m.) and mice were transcardially perfused 2 h later. Representative brain sections stained with DAPI (blue) showing OVA-A647 (red) influx into the brain parenchyma of ligated and sham-operated mice. Scale bar, 5 mm and 2 mm (inset). **b**, Quantification of OVA-A647 area fraction (%) in brain sections showed a significant decrease in the ligation group. **c**, Representative sections of dCLNs stained with DAPI (blue) and for LYVE-1 (green), showing OVA-A647 (red) coverage in the ligation and sham-operated groups. Scale bar, 200 μ m. **d**, Quantification of OVA-A647 area fraction (%) in the dCLNs showed a significant decrease in the ligation group. Data in **b** and **d** are mean \pm s.e.m., $n = 8$ per group, two-tailed Mann–Whitney *U*-test; data in **a–d** were pooled from two independent experiments and are representative of three independent experiments. **e**, Wild-type (WT) and *Prox1*^{+/-} mice (2–3 months old) were injected with 5 μ l of OVA-A647 into the CSF (i.c.m.) and transcardially perfused 2 h later. **f**, Representative brain sections stained with DAPI (blue) showing OVA-A647 (red) influx into the brain parenchyma of *Prox1*^{+/-} and wild-type mice. Scale bar, 5 mm. **g**, Quantification of OVA-A647 area fraction (%) in brain sections showed a significant decrease in *Prox1*^{+/-} mice. **h**, Representative sections of dCLNs stained with DAPI (blue) and for LYVE-1 (green), showing OVA-A647 (red) coverage in the dCLNs of *Prox1*^{+/-} and wild-type mice. Scale bar, 500 μ m. **i**, Quantification of OVA-A647 area fraction (%) in the dCLNs showed a significant decrease in *Prox1*^{+/-} mice. Data in **g** and **i** are mean \pm s.e.m., $n = 15$ wild-type mice, $n = 12$ *Prox1*^{+/-} mice, two-tailed Mann–Whitney *U*-test; data in **e–i** were pooled from two independent experiments. **j**, Rate of brain paravascular influx of the contrast agent Gd, injected i.c.m. at 1, 10 or 25 mM (in saline), was assessed in adult mice (3 months old) by T1-weighted MRI. **k**, Representative MRI images obtained using Lymph4D

software showing brain signal intensity for different concentrations of injected Gd. Scale bar, 3 mm. Experiments in **j** and **k** were performed once. **l**, Adult mice were subjected to meningeal lymphatic ablation by visudyne photoconversion. One week later, T1-weighted MRI acquisition was performed after i.c.m. injection of 5 μ l of Gd (25 mM in saline). Using the Lymph4D software, it was possible to measure the rate of contrast agent influx into the delineated brain cortical region of mice from both groups. Scale bar, 3 mm. Images in sequence 2 and subsequent were obtained by subtraction of sequence 1. **m**, Quantification of the signal intensity gain (relative to sequence 1) in the brain cortex revealed a significant decrease in the visudyne with photoconversion group, when compared to vehicle with photoconversion. **n**, **o**, Coronal sections of the brain of vehicle- or visudyne-treated mice ($n = 4$ per group) were aligned and stacked into 2D colour maps (concatenated from 16 MRI sequences) showing contrast of Gd signal intensity (**n**) and isotropic diffusion coefficient (**o**). Scale bars, 3 mm. **p**, Area fraction quantification of high, medium and low values of isotropic diffusion coefficient in the four 2D stacks, in visudyne relative to vehicle. Data in **m** and **p** are mean \pm s.e.m., $n = 4$ per group, repeated-measures two-way ANOVA with Bonferroni's post hoc test (**m**); one-way ANOVA with Bonferroni's post hoc test (**p**). **l–p**, Data are representative of two independent experiments. **q**, Representative confocal images of DAPI (blue) and aquaporin 4 (AQP4, green) staining and OVA-A647 (red) levels in brain sections from vehicle- and visudyne-treated mice. Scale bar, 500 μ m. **r**, Quantification of area fraction (%) of AQP4 in the brains of mice treated with vehicle or visudyne shows that there are no differences between groups. **s**, Images showing representative staining for AQP4⁺ astrocytic endfeet (red) and CD31⁺ blood vessels (green) in the brain cortex of mice from vehicle and visudyne groups. Scale bar, 50 μ m. **t–v**, No changes were observed in the area of AQP4⁺ astrocytic endfeet (**t**) and CD31⁺ blood vessels (**u**) or in the ratio between area of AQP4⁺ and of CD31⁺ (**v**). Data in **r**, **t–v** are mean \pm s.e.m., $n = 7$ per group, two-tailed Mann–Whitney *U*-test; data in **q–v** were pooled from two independent experiments and representative of three independent experiments.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Ablation of meningeal lymphatic vessels impairs efflux of macromolecules from the brain. **a**, Seven days after meningeal lymphatic ablation, 1 μ l of fluorescent OVA-A647 (0.5 mg ml⁻¹ in artificial CSF) was stereotactically injected (coordinates from bregma, AP, +1.5 mm; ML, -1.5 mm; DV, +2.5 mm) into the brain parenchyma. **b**, Representative brain sections rostral and caudal to the injection site, stained for glial fibrillary acidic protein (GFAP, in green), demonstrating OVA-A647 (red) coverage of the brain parenchyma in the visudyne with photoconversion group and the control groups. Scale bar, 5 mm. **c**, Quantification of OVA-A647 area fraction (%) in the injected brain hemisphere showing a significantly higher level in the visudyne with photoconversion group compared to both control groups. Data are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test. **d**, Seven days after meningeal lymphatic ablation, 1 μ l of fluorescent amyloid- β_{42} (A β_{42})-HiLyte647 (0.05 μ g ml⁻¹ in artificial CSF) was stereotactically injected (coordinates from bregma, AP, +1.5 mm; ML, -1.5 mm; DV, +2.5 mm) into the brain parenchyma. **e**, Representative brain sections rostral and caudal to the injection site, stained for GFAP

(green), demonstrating A β_{42} -HiLyte647 (red) coverage of the brain parenchyma in the visudyne with photoconversion group and the control groups. Scale bar, 5 mm. **f**, Quantification of A β_{42} -HiLyte647 area fraction (%) in the injected brain hemisphere showing a significantly higher level in the visudyne with photoconversion group compared to both control groups. Data are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test. **g**, Seven days after meningeal lymphatic ablation, 1 μ l of fluorescent LDL-BODIPY FL (0.1 mg ml⁻¹ in artificial CSF) was stereotactically injected (coordinates from bregma, AP, +1.5 mm; ML, -1.5 mm; DV, +2.5 mm) into the brain parenchyma. **h**, Representative brain sections rostral and caudal to the injection site, stained for GFAP (red), demonstrating LDL-BODIPY FL (green) coverage of the brain parenchyma in the visudyne with photoconversion group and the control groups. Scale bar, 5 mm. **i**, Quantification of LDL-BODIPY FL area fraction (%) in the injected brain hemisphere showing a significantly higher level in the visudyne with photoconversion group compared to both control groups. Data are mean \pm s.e.m., $n = 6$ per group, one-way ANOVA with Bonferroni's post hoc test.



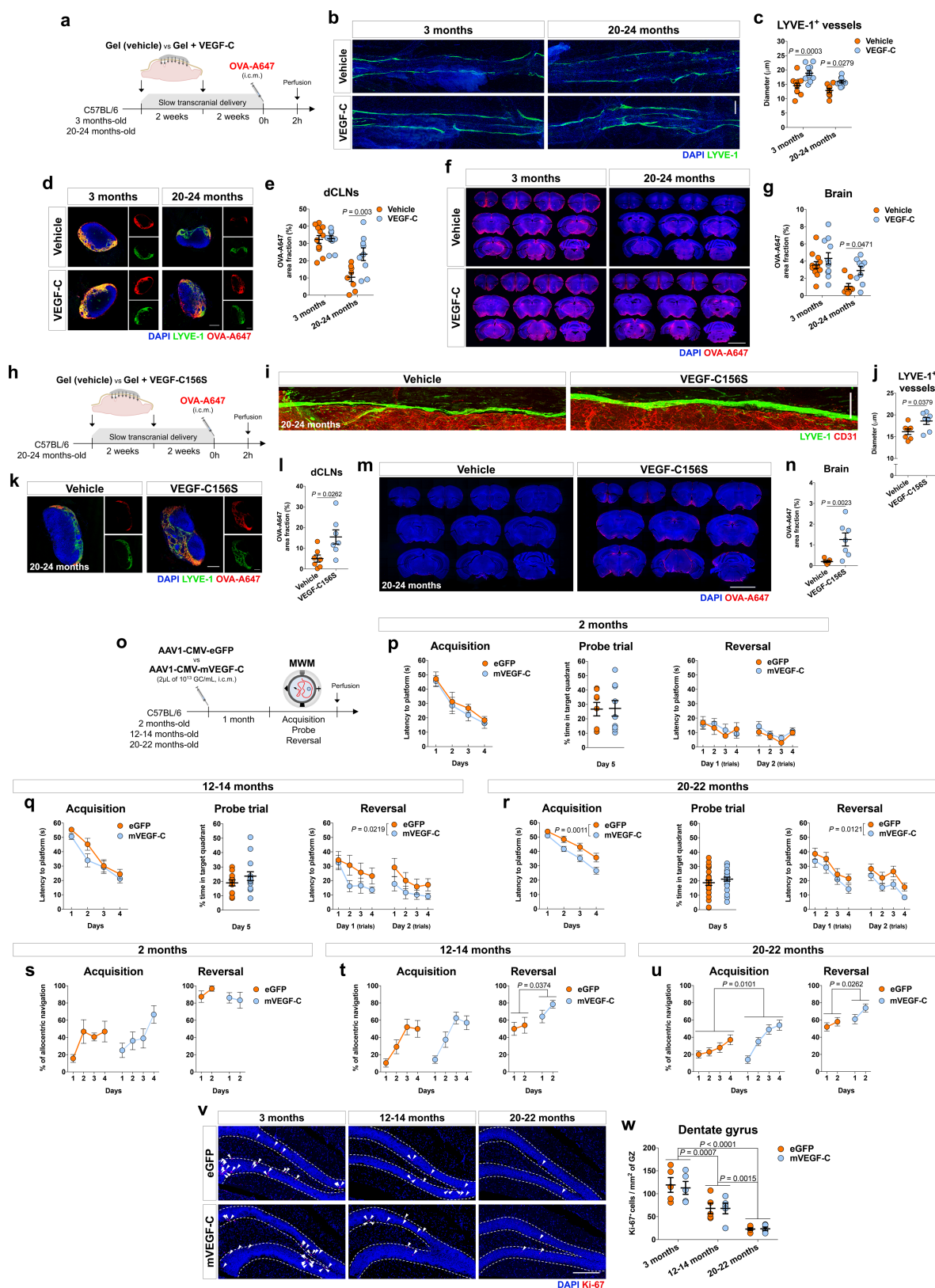
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Behavioural assessment and hippocampal RNA-seq analysis after impairing meningeal lymphatic function. **a, b**, No differences in total distance travelled (**a**) and time in centre of the open field arena (**b**) were observed between vehicle with photoconversion, visudyne and visudyne with photoconversion groups. Data are mean \pm s.e.m., $n = 9$ per group, one-way ANOVA with Bonferroni's post hoc test. **c, d**, Performance of mice from the three groups was also identical both during the training session (**c**) and during the novel location task (**d**) of the NLR paradigm. Data are mean \pm s.e.m., $n = 9$ per group, two-way ANOVA with Bonferroni's post hoc test. **e, f**, Performance of mice in the CFC paradigm showed no differences between groups in the context test (**e**), however, there was a statistically significant difference in the cued test (**f**). Data are mean \pm s.e.m., $n = 9$ per group, one-way ANOVA with Bonferroni's post hoc test. **g**, The cognitive performance of adult mice was assessed in the MWM test, one week after sham surgery or surgical ligation of the lymphatic vessels afferent to the dCLNs. **h**, Ligated mice showed a significant increase in the latency to platform during acquisition compared to sham-operated mice. **i, j**, No significant differences between groups were observed in the percentage of time spent in the target quadrant in the probe trial (**i**) or in latency to platform in the reversal (**j**). Data are mean \pm s.e.m., $n = 8$ sham-operated mice, $n = 9$ mice with ligation; repeated-measures two-way ANOVA with Bonferroni's post hoc test (**h, j**), two-tailed Mann-Whitney U -test (**i**). **k**, Vehicle or visudyne injection experiments with photoconversion were performed twice within a two-week interval. Total RNA was extracted from the hippocampus of mice from both groups and sequenced (RNA-seq). RNA-seq principal component (PC) analysis did not show a differential clustering of samples from vehicle and visudyne groups. **l**, Heat map showing relative expression levels of genes in vehicle with photoconversion and in visudyne with photoconversion samples. **m**, After meningeal lymphatic

ablation (twice within a two-week interval) and MWM performance, total RNA was extracted from the hippocampus of mice from vehicle with photoconversion or visudyne with photoconversion groups and sequenced. RNA-seq principal component analysis demonstrating a differential clustering of samples from vehicle and visudyne groups. A total of 2,138 genes were downregulated and 1,599 genes were upregulated in the hippocampus after meningeal lymphatic ablation and MWM performance. **n**, Heat map showing relative expression levels of genes in vehicle with photoconversion and in visudyne with photoconversion samples. Colour scale bar values represent standardized rlog-transformed values across samples (**l, n**). **o**, Neurological disease, neuronal activity- and synaptic plasticity-related GO and KEGG terms are enriched upon visudyne treatment, as measured by the $-\log_{10}(\text{adjusted } P \text{ value})$. **p**, GO and KEGG terms related to metabolite generation and processing, glycolysis and mitochondrial respiration and oxidative stress were enriched, as measured by the $-\log_{10}(\text{adjusted } P \text{ value})$, upon visudyne treatment and MWM performance. **q, r**, Heat map showing relative expression levels of genes involved in two of the significantly altered GO terms related to excitatory synapse (**q**) and learning or memory (**r**). **s–v**, Heat maps showing relative expression levels of genes involved in four of the significantly altered GO terms related to NADH dehydrogenase complex (**s**), generation of precursor metabolites and energy (**t**), cellular response to oxidative stress (**u**) and cellular response to nitrogen compound (**v**). Datasets in **k–v** all consist of $n = 5$ per group; **k, m**, P values were corrected for multiple hypothesis testing with the Benjamini–Hochberg false-discovery rate procedure; in **l, n–v**, functional enrichment of differential expressed genes was performed using gene sets from GO and KEGG and determined with Fisher's exact test. Colour scale bar values in **n, q–v** represent standardized rlog-transformed values across samples.

Extended Data Fig. 6 | Characterization of meningeal lymphatics in young and old mice and improvement of lymphatic function by viral-mediated expression of mVEGF-C. **a**, OVA-A647 was injected into the CSF (i.c.m.) of young-adult (3 months of age) and old (20–24 months of age) mice. Representative brain sections stained with DAPI (blue) showing degree of OVA-A647 (red) influx into the parenchyma. Scale bars, 5 mm and 2 mm (inset). **b**, Quantification of OVA-A647 area fraction (%) in brain sections. Data are mean \pm s.e.m., $n = 6$ mice in 3 months, $n = 8$ mice in 20–24 months, two-tailed Mann–Whitney *U*-test; representative of two independent experiments. **c**, Representative images of DAPI (blue) and LYVE-1 (green) staining in meningeal whole-mounts of young-adult (2 months old) and old (20–24 months old) male and female mice. Scale bar, 1 mm. **d**, Measurement of LYVE-1⁺ vessel diameter and area fraction showed a significant decrease in both parameters in old mice, when compared to young-adults, in both females and males. **e**, Representative images of DAPI (blue) and LYVE-1 (green) staining in dCLNs 2 h after injection of OVA-A594 (red) into the CSF of young-adult and old mice from both genders. Scale bar, 200 μ m. **f**, Quantification of OVA-A594 area fraction (%) in the dCLNs of mice from different ages and genders mice. Data in **d**, **f** are mean \pm s.e.m., $n = 9$ per group at 2 months, $n = 7$ per group at 20–24 months for male and female mice, two-way ANOVA with Bonferroni's post hoc test; data were pooled from two independent experiments. **g**, Representative dot and contour plots showing the gating strategy used to isolate meningeal LECs by FACS from the meninges of young-adult and old mice. $n = 3$ per group, pooled from two independent experiments. **h**, Adult mice were injected i.c.m. with 2 μ l of AAV1-CMV-eGFP (eGFP) or AAV1-CMV-mVEGF-C (mVEGF-C), both at 10^{13} GC ml⁻¹, and transcardially perfused with saline 2 or 4 weeks later. **i**, Representative brain coronal sections of mice showing eGFP⁺ infected cells (green) in the pia mater, surrounding the GFAP⁺ glia limitans (red) of the brain parenchyma at 2 and 4 weeks post injection. Scale bars, 5 mm and 200 μ m (inset). **j**, Representative insets from meningeal whole-mounts stained

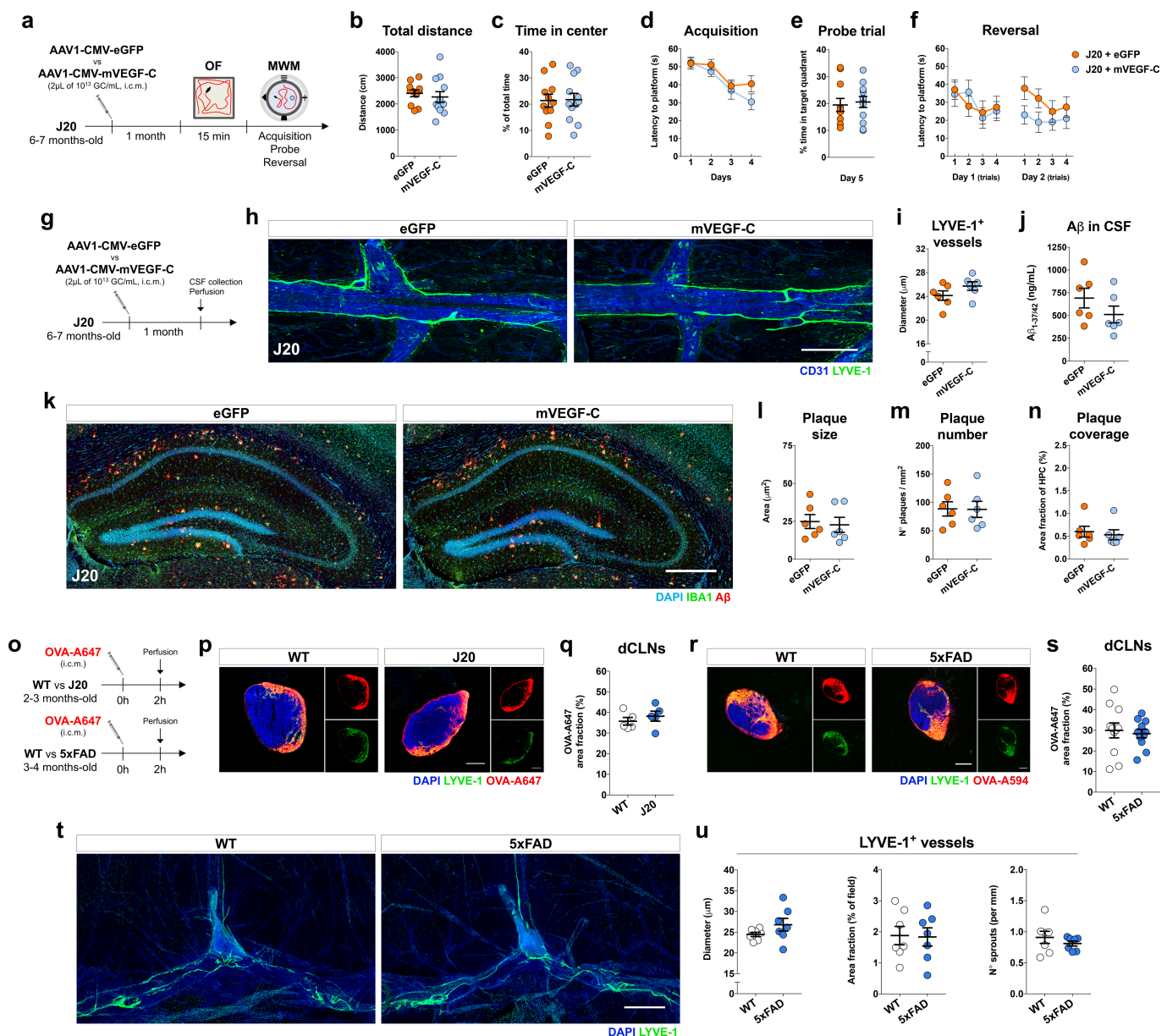
for CD31 (blue), eGFP (green) and LYVE-1 (red). Scale bar, 200 μ m. Green cells are observed in the cerebellar meninges, pineal gland and transverse sinus in the eGFP group at 2 and 4 weeks, but not in the same regions of the meninges in the mVEGF-C group. **k**, Representative images of LYVE-1⁺ lymphatic vessels (red) and LYVE-1⁻CD31⁺ blood vessels (blue) in the superior sagittal sinus of mice treated with either eGFP or mVEGF-C for 2 or 4 weeks. Scale bar, 200 μ m. **l**, **m**, Mice treated with AAV1 expressing mVEGF-C presented a significant increase in lymphatic vessel diameter (**l**), but not in coverage by blood vessels (**m**). Data in **l** and **m** are mean \pm s.e.m., $n = 4$ per group at 2 weeks, $n = 3$ per group at 4 weeks; two-way ANOVA with Bonferroni's post hoc test; data in **h–m** are representative of two independent experiments. **n**, Representative images of blood flow (mm s⁻¹) and arterial and venous blood oxygenation (percentage of sO₂) readings obtained by photoacoustic imaging of brain/meningeal vasculature of old mice (20–22 months old) treated for one month with eGFP or mVEGF-C virus (both at 10^{13} GC ml⁻¹). **o**, **p**, The different treatments did not affect blood flow (**o**) or blood oxygenation (**p**) in the brain/meninges of old mice. Data are mean \pm s.e.m., $n = 5$ per group; two-tailed Mann–Whitney *U*-test (**o**), two-way ANOVA with Bonferroni's post hoc test (**p**); data obtained from a single experiment. **q**, Old mice (20–22 months old) were injected i.c.m. with 2 μ l of viral vectors expressing eGFP or mVEGF-C. One month later, T1-weighted MRI acquisition was performed after i.c.m. injection of 5 μ l of Gd (25 mM in saline). Using the Lymph4D software, it was possible to measure the rate of contrast agent influx into the delineated brain hippocampal region of mice from both groups. Scale bar, 3 mm. Images in sequence 2 and subsequent were obtained by subtraction of sequence 1. **r**, Quantification of the signal intensity gain (relative to sequence 1) in the hippocampus revealed a significant increase in the mVEGF-C group compared to eGFP. Data are mean \pm s.e.m., $n = 4$ per group; repeated-measures two-way ANOVA with Bonferroni's post hoc test; data were pooled from two independent experiments.



Extended Data Fig. 7 | See next page for caption.

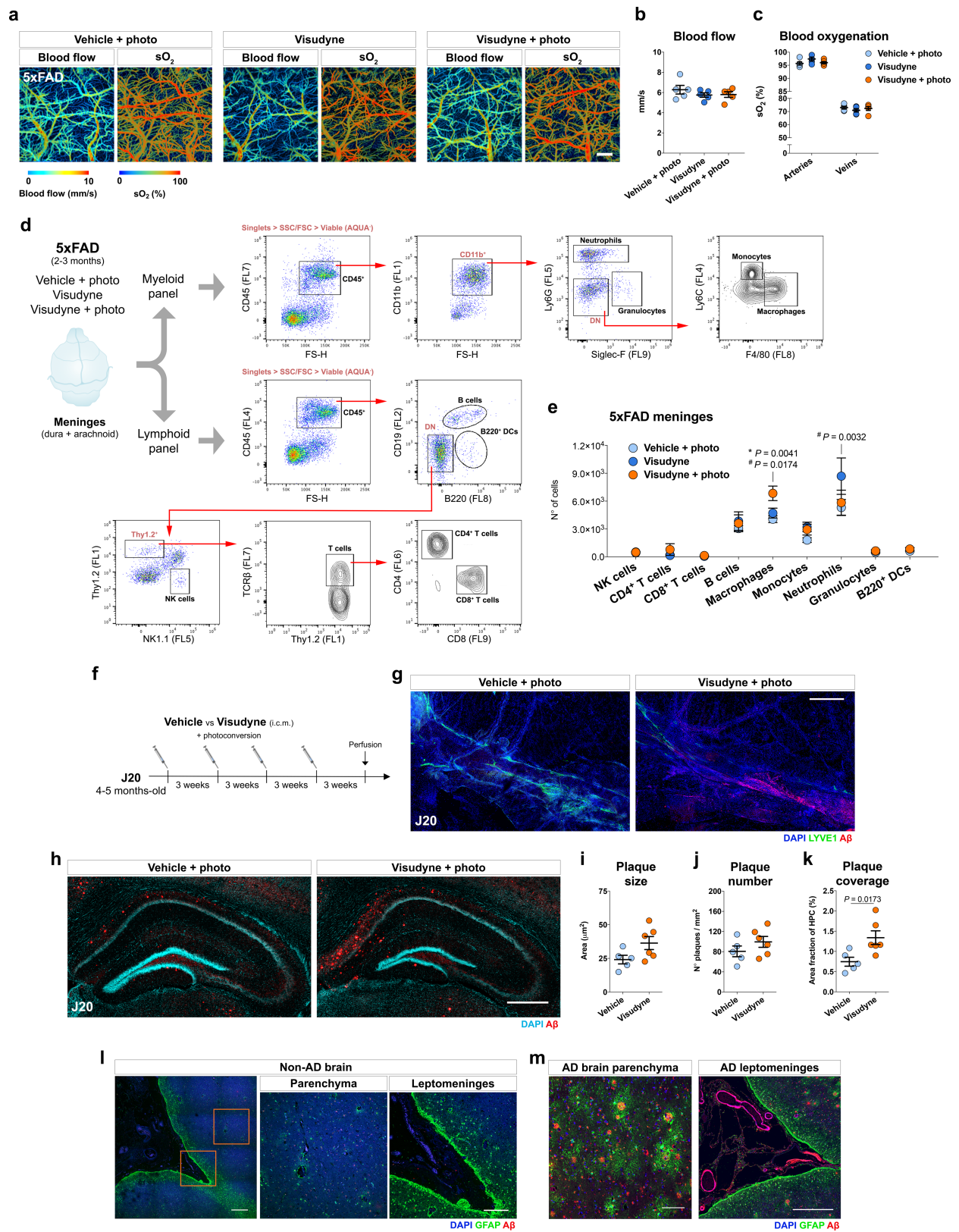
Extended Data Fig. 7 | Treatment with VEGF-C ameliorates meningeal lymphatic function, brain perfusion by CSF macromolecules and cognitive performance in old mice. **a**, Hydrogel alone (vehicle) or containing recombinant human VEGF-C (200 ng ml^{-1}) was applied on top of a thinned skull surface of adult mice (3 months old) and old mice (20–24 months old). Gels were reapplied two weeks later. Four weeks after the initial treatment, $5 \mu\text{l}$ of OVA-A647 (in artificial CSF) was injected into the CSF (i.c.m.) and mice were transcardially perfused 2 h later. **b**, Representative images of DAPI (blue) staining and LYVE-1⁺ vessels (in green) in the superior sagittal sinus after transcranial delivery of VEGF-C. Scale bar, $50 \mu\text{m}$. **c**, Treatment with VEGF-C resulted in a significant increase in lymphatic vessel diameter in the superior sagittal sinus in both adult and old mice. **d**, Representative sections of dCLNs stained with DAPI (blue) and for LYVE-1 (green) showing drained OVA-A647 (red). Scale bars, $200 \mu\text{m}$. **e**, Quantification of OVA-A647 (red) area fraction (%) in the dCLNs showed increased drainage in old mice treated with VEGF-C compared to vehicle-treated age-matched mice. **f**, Representative brain sections stained with DAPI (blue) showing OVA-A647 (red) influx into the brain parenchyma. Scale bar, 5 mm. **g**, Influx of OVA-A647 into the brain parenchyma of old mice was significantly increased after transcranial delivery of VEGF-C. Data in **c**, **e** and **g** are mean \pm s.e.m., $n = 12$ mice treated with vehicle at 3 months, $n = 11$ mice treated with VEGF-C at 3 months, $n = 8$ mice treated with vehicle at 20–24 months and $n = 9$ mice treated with VEGF-C at 20–24 months; two-way ANOVA with Bonferroni's post hoc test; data in **a–g** were pooled from two independent experiments. **h**, Hydrogel alone (vehicle) or containing recombinant human VEGF-C156S (200 ng ml^{-1}) was applied on top of a thinned skull surface of old mice. Gels were reapplied two weeks later. **i**, Whole-mounts of brain meninges were stained for LYVE-1 (green) and CD31 (red). Images show insets of lymphatic vessels near the superior sagittal sinus. Scale bar, $100 \mu\text{m}$. **j**, Old mice that received VEGF-C156S treatment showed increased diameter of LYVE-1⁺ vessels in the superior sagittal sinus. **k**, Representative sections of dCLNs stained with DAPI (blue) and for LYVE-1 (green) showing levels of OVA-A647 (red) drained from the CSF. Scale bar, $200 \mu\text{m}$. **l**, Quantification of OVA-A647 area fraction (%) in the dCLNs showed a significant increase in VEGF-C156S group compared to vehicle. **m**, Representative images of OVA-A647 (red) in brain sections also stained with DAPI (blue). Scale bar, 5 mm. **n**, Quantification of OVA-A647 area fraction (%) in brain sections showed a significant increase in brain influx of the tracer in old mice treated with VEGF-C156S. Data in **j**, **l** and **n** are mean \pm s.e.m., $n = 7$ mice per group; two-tailed Mann–Whitney *U*-test; data in **h–n** were pooled from two independent experiments. **o**, Young-adult (2 months), middle-aged (12–14 months)

or old (20–22 months) mice were injected with viral vectors expressing eGFP or mVEGF-C. One month after injection, learning and memory was assessed using the MWM test. **p**, Injection of mVEGF-C virus in young-adult mice did not alter their performance in the acquisition, probe trial or reversal of the MWM. Data are mean \pm s.e.m., $n = 8$ mice treated with eGFP and $n = 9$ mice treated with mVEGF-C; repeated-measures two-way ANOVA with Bonferroni's post hoc test was used in the acquisition and reversal; two-tailed Mann–Whitney *U*-test was used in the probe trial; data were obtained in a single experiment. **q**, Injection of mVEGF-C virus in middle-aged mice did not alter their performance in the acquisition and in the probe trial, but significantly improved their performance in the reversal. Data are mean \pm s.e.m., $n = 12$ mice treated with eGFP and $n = 14$ mice treated with mVEGF-C; repeated-measures two-way ANOVA with Bonferroni's post hoc test was used in the acquisition and reversal, two-tailed Mann–Whitney *U*-test was used in the probe trial; data were pooled from two independent experiments. **r**, Injection of mVEGF-C virus in old mice did not alter their performance in the probe trial, but significantly improved their performance in the acquisition and in the reversal. Data are mean \pm s.e.m., $n = 25$ mice treated with eGFP and $n = 25$ mice treated with mVEGF-C; repeated-measures two-way ANOVA with Bonferroni's post hoc test was used in the acquisition and reversal; two-tailed Mann–Whitney *U*-test was used in the probe trial; data were pooled from three independent experiments. **s**, Treatment of young-adult mice with mVEGF-C did not affect the percentage of allocentric navigation strategies used in the MWM. **t**, **u**, The percentage of allocentric navigation strategies was significantly higher in middle-aged mice treated with mVEGF-C during the reversal (**t**) and in old mice treated with mVEGF-C during the acquisition and reversal (**u**) compared to age-matched eGFP-treated mice. Data in **s–u** are mean \pm s.e.m., $n = 8$ mice treated with eGFP and $n = 9$ mice treated with mVEGF-C at 2 months (**s**), $n = 12$ mice treated with eGFP and $n = 14$ mice treated with mVEGF-C at 12–14 (**t**), $n = 25$ per group at 20–22 months (**u**); **s–u**, repeated-measures two-way ANOVA with Bonferroni's post hoc test; data were obtained from a single experiment (**s**), pooled from two (**t**) and three (**u**) independent experiments. **v**, Insets of the hippocampal dentate gyrus (granular zone (GZ)), stained with DAPI (blue) and for Ki-67 (in red), in mice injected with viral vectors expressing eGFP or mVEGF-C at 2, 12–14 and 20–22 months. Scale bar, $200 \mu\text{m}$. **w**, Ageing induced a significant decrease in Ki-67⁺ proliferating cells in the dentate gyrus. Expression of mVEGF-C in the meninges at the analysed ages did not affect the number of Ki-67⁺ cells in the dentate gyrus. Data are mean \pm s.e.m., $n = 5$ per group, two-way ANOVA with Bonferroni's post hoc test.



Extended Data Fig. 8 | Expression of mVEGF-C in the meninges of J20 mice does not ameliorate lymphatic drainage or brain amyloid pathology. **a**, J20 mice were injected i.c.m. with 2 μ L of AAV1-CMV-eGFP or AAV1-CMV-mVEGF-C (10^{13} GC mL⁻¹) at 6–7 months. One month after injection, the mice were tested in the open field (OF) and in the MWM. **b**, **c**, Total distance and percentage of time in the centre of the open field arena was not ameliorated by treatment of J20 mice with mVEGF-C. **d–f**, No statistically significant differences were observed in the acquisition (**d**), in the probe trial (**e**) or in the reversal (**f**) of the MWM test after one month of mVEGF-C treatment. Data in **b–f** are mean \pm s.e.m., $n = 11$ mice treated with eGFP, $n = 12$ mice treated with mVEGF-C; two-tailed Mann–Whitney *U*-test (**b**, **c**, **e**), repeated-measures two-way ANOVA with Bonferroni's post hoc test (**d**, **f**); data were from a single experiment. **g**, J20 mice were treated with eGFP or mVEGF-C and, one month later, CSF, meninges and brain were collected for analysis. **h**, Representative images of DAPI (blue) and LYVE-1⁺ lymphatic vessels (green) in the superior sagittal sinus of mice treated with either eGFP or mVEGF-C. Scale bar, 500 μ m. **i**, AAV1-mediated expression of mVEGF-C did not affect meningeal lymphatic vessel diameter. **j**, Levels of amyloid- β in the CSF measured by ELISA remained unaltered after mVEGF-C treatment. **k**, Representative images of dorsal hippocampi of J20 mice treated with eGFP or mVEGF-C stained with DAPI (cyan) and for IBA1 (green) and amyloid- β (red). Scale bar, 500 μ m. **l–n**, No changes were observed in amyloid plaque size (**l**), number (**m**) or coverage (**n**) between

the groups. Data in **i**, **j**, **l–n** are mean \pm s.e.m., $n = 6$ per group, two-tailed Mann–Whitney *U*-test; data in **g–n** obtained from a single experiment. **o**, J20 mice (2–3 months old), 5xFAD mice (3–4 months old) and respective age-matched wild-type littermate controls were injected with fluorescent OVA-A647 (i.c.m.) in order to measure drainage into the dCLNs. **p**, Representative images of DAPI (blue) and LYVE-1 (green) staining in dCLNs of wild-type and J20 mice 2 h after injection of OVA-A647 (red). Scale bar, 200 μ m. **q**, Quantification of OVA-A647 area fraction (%) in the dCLNs shows equal levels of tracer in mice from both genotypes. Data are mean \pm s.e.m., $n = 5$ per group, two-tailed Mann–Whitney *U*-test, representative of two independent experiments. **r**, Representative images of DAPI (blue) and LYVE-1 (green) staining in dCLNs of wild-type and 5xFAD mice 2 h after injection of OVA-A594 (red). Scale bars, 200 μ m. **s**, Quantification of OVA-A594 area fraction (%) in the dCLNs shows equal levels of tracer in mice from both genotypes. Data are mean \pm s.e.m., $n = 11$ per group, two-tailed Mann–Whitney *U*-test, data were pooled from two independent experiments. **t**, Representative images of DAPI (blue) and LYVE-1 (green) staining in meningeal whole-mounts of wild-type and 5xFAD mice at 3–4 months. Scale bar, 1 mm. **u**, Measurement of LYVE-1⁺ vessel diameter, area fraction and number of sprouts (per mm of vessel) showed no differences between genotypes. Data are mean \pm s.e.m., $n = 7$ per group, two-tailed Mann–Whitney *U*-test; data were pooled from two independent experiments.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Meningeal lymphatic ablation in transgenic mice with Alzheimer's disease worsens amyloid pathology without affecting blood vessel function. **a**, Representative images of blood flow (mm s^{-1}) and arterial and venous blood oxygenation (percentage of sO_2) readings obtained by photoacoustic imaging of brain/meningeal vasculature of 5xFAD mice one week after vehicle with photoconversion, visudyne or visudyne with photoconversion. **b, c**, The different treatments did not affect blood flow (**b**) or blood oxygenation (**c**) in the brain/meninges of 5xFAD mice. Data are mean \pm s.e.m., $n = 5$ per group, one-way ANOVA with Bonferroni's post hoc test (**b**), two-way ANOVA with Bonferroni's post hoc test (**c**), data were from a single experiment. **d**, Representative flow cytometry dot and contour plots showing the gating strategies used to determine the frequency of specific immune cell populations, using a myeloid or lymphoid panel of markers, in the meninges of 5xFAD after prolonged (1.5 months) meningeal lymphatic ablation. **e**, Analysis of specific immune cell populations in the meninges of 5xFAD mice from the different groups showed a significant increase in macrophages in the visudyne with photoconversion group compared to the control groups. A significant increase in neutrophils was observed in visudyne group, but not in vehicle with photoconversion group compared to visudyne with photoconversion group. Data are mean \pm s.e.m., $n = 5$ per group; two-way ANOVA with Holm–Sidak's post hoc test, *versus vehicle with photoconversion, #versus visudyne; data obtained from a single

experiment. **f**, J20 mice aged 4–5 months were subjected to meningeal lymphatic ablation by injection (i.c.m.) of visudyne or vehicle as a control, followed by a photoconversion step. This procedure was repeated every three weeks, for a total of three months, to achieve prolonged meningeal lymphatic ablation. **g**, Staining with DAPI (blue) and for LYVE-1 (green) and amyloid- β (red) in meningeal whole-mounts of J20 mice showing marked amyloid deposition in mice from the visudyne group. Scale bar, $500 \mu\text{m}$. **h**, Representative brain sections of J20 mice at 7–8 months of age stained with DAPI (cyan) and for amyloid- β (red) showing degree of amyloid deposition after meningeal lymphatic ablation. Scale bar, $500 \mu\text{m}$. **i–k**, Quantification of amyloid plaque size (**i**), number (**j**) and coverage (**k**) in the dorsal hippocampus of J20 mice showed a statistically significant increase in coverage in the visudyne group compared to vehicle. Data in **i–k** are mean \pm s.e.m., $n = 5$ vehicle-treated mice, $n = 6$ visudyne-treated mice; two-tailed Mann–Whitney U -test; experiments in **f–k** were performed once. **l, m**, Sections of human brain cortex, containing meningeal layers (leptomeninges) attached, from the brain of a control (**l**) (scale bars, $500 \mu\text{m}$ and $200 \mu\text{m}$ (inset)) and the brain of a patient with Alzheimer's disease (**m**) (scale bars, $100 \mu\text{m}$ (left) and $500 \mu\text{m}$ (right)) were stained with DAPI (blue), for the astrocyte marker GFAP (green) and for amyloid- β (red). Data in **l** and **m** are from $n = 8$ controls and $n = 9$ patients with Alzheimer's disease and are representative of two independent experiments.

Extended Data Table 1 | Demographic data of patients with and without Alzheimer's disease

	Age (years)	Gender	Diagnosis criteria	Pathological score
AD	62	F	Intermediate probability*	A2, B3, C2-3; CAA
	64	M	Possible	CERAD C; BB I/II; CAA
	72	M	High probability*	A3, B3, C3; CAA
	76	F	High probability*	A3, B3, C3; CAA
	79	M	High probability*	A3, B3, C3; CAA
	83	F	High probability*	A3, B3, C3; CAA
	83	M	Intermediate probability*	A2, B2, C2
	88	F	High probability*	A3, B3, C3; CAA
	95	F	Definitive	CERAD C; BB V/VI; CAA
Mean \pm s.e.m.	78 \pm 3.6			
	Age (years)	Gender	Cause of death	
Non-AD	63	F	Multi-organ failure after motor vehicle accident	
	63	M	Acute myocardial infarct	
	64	M	Bilateral pulmonary emboli	
	65	F	Decompensated ischemic cardiomyopathy	
	70	M	Bronchopneumonia	
	73	F	Septicemia	
	80	F	Bronchopneumonia	
	91	F	Cardiovascular atherosclerotic disease	
Mean \pm s.e.m.	71.1 \pm 3.5			

AD, Alzheimer's disease; BB, Braak and Braak stage; CAA, cerebral amyloid angiopathy; CERAD, Consortium to Establish a Registry for Alzheimer's disease.

*New criteria for diagnosis following the guidelines of NIA-AA based on ABC (Amyloid, Braak, CERAD) score.

Cyclin-dependent kinase 12 is a drug target for visceral leishmaniasis

Susan Wyllie¹, Michael Thomas¹, Stephen Patterson¹, Sabrinia Crouch², Manu De Rycker¹, Rhiannon Lowe³, Stephanie Gresham³, Michael D. Urbaniak^{1,4}, Thomas D. Otto^{5,9}, Laste Stojanovski¹, Frederick R. C. Simeons¹, Sujatha Manthri¹, Lorna M. MacLean¹, Fabio Zuccotto¹, Nadine Homeyer¹, Hannah Pflaumer⁶, Markus Boesche⁶, Lalitha Sastry¹, Paul Connolly⁷, Sebastian Albrecht¹, Matt Berriman⁵, Gerard Drewes⁶, David W. Gray¹, Sonja Ghidelli-Disse⁶, Susan Dixon⁸, Jose M. Fiandor², Paul G. Wyatt¹, Michael A. J. Ferguson¹, Alan H. Fairlamb¹, Timothy J. Miles^{2*}, Kevin D. Read^{1*} & Ian H. Gilbert^{1*}

Visceral leishmaniasis causes considerable mortality and morbidity in many parts of the world. There is an urgent need for the development of new, effective treatments for this disease. Here we describe the development of an anti-leishmanial drug-like chemical series based on a pyrazolopyrimidine scaffold. The leading compound from this series (7, DDD853651/GSK3186899) is efficacious in a mouse model of visceral leishmaniasis, has suitable physicochemical, pharmacokinetic and toxicological properties for further development, and has been declared a preclinical candidate. Detailed mode-of-action studies indicate that compounds from this series act principally by inhibiting the parasite cdc-2-related kinase 12 (CRK12), thus defining a druggable target for visceral leishmaniasis.

Leishmania parasites cause a wide spectrum of human infections ranging from the life-threatening visceral disease to disfiguring mucosal and cutaneous forms. *Leishmania* spp. are obligate intracellular parasites of the vertebrate reticuloendothelial system, where they multiply as amastigotes within macrophage phagolysosomes; transmission is by blood-sucking sandflies, in which they proliferate as extracellular promastigotes.

Visceral leishmaniasis, resulting from infection with *Leishmania donovani* and *L. infantum*, causes 20,000–40,000 deaths annually according to the WHO (World Health Organization; <http://www.who.int/leishmaniasis/en/>), of which approximately 60% occur in India, Bangladesh and Nepal¹. In 95% of cases, death can be prevented by timely and appropriate drug therapy². However, current treatment options are far from ideal, with outcomes depending on several factors including geographical location, the immune status and other comorbidities of the patient, and the disease classification. None of the current front-line treatments for visceral leishmaniasis—amphotericin B (liposomal or deoxycholate formulations), miltefosine, paromomycin and antimonials—is ideal for use in resource-poor settings, owing to issues such as teratogenicity, cost, resistance and/or clinical relapse, prolonged treatment regimens and parenteral administration^{3–5}. Thus, there is an urgent need for new treatment options for visceral leishmaniasis, particularly oral drugs. Unfortunately, there are currently, to our knowledge, no new therapeutics in clinical development and only a few in preclinical development. There is a paucity of well-validated molecular drug targets in *Leishmania*, and the molecular targets of the current clinical molecules are unknown. Recent studies⁶ identified the proteasome as a promising therapeutic target for the treatment of visceral leishmaniasis as well as other kinetoplastid infections, and this currently represents the most robustly validated drug target that has been reported in these parasites. Furthermore, whole-cell (phenotypic) screening programs have been hindered by extremely low hit rates⁷.

Here, we report the discovery of a promising anti-leishmanial compound with a new mechanism of action.

Discovery

Previously, we reported the identification of a diaminothiazole series from a compound screen against *Trypanosoma brucei* GSK3 kinase (*TbGSK3*)⁸. During compound optimization, it became clear that the anti-trypanosomal activity of this series was driven, at least in part, by off-target activity. The early compounds showed activity against *T. brucei* bloodstream trypanosomes in viability assays, but showed little activity against *L. donovani* axenic amastigotes (for example, compound 1). Modification of the core structure, while retaining the functions of the hydrogen bond donor and acceptor, gave a bicyclic compound series (Fig. 1), one of which (compound 2) showed very weak activity against *L. donovani* axenic amastigotes, but was inactive against the more clinically relevant intra-macrophage amastigotes. Appending a sulfonamide to the cyclohexyl ring resulted in compound 3, which was active against *L. donovani* amastigotes in both the axenic and intra-macrophage assays^{9,10} and selectively active against *L. donovani* compared to the THP-1 mammalian host cells used in the assay. Replacement of the *iso*-butyl substituent on the pyrazole ring with an aromatic substituent and the benzyl group on the sulfonamide with a trifluoropropyl substituent resulted in compound 4, which had marginally more activity. Notably, this compound demonstrated more than 70% parasite reduction in a mouse model of visceral leishmaniasis when dosed orally, providing proof of concept in an animal model for this series. Replacing the pyridyl group with a 2-methoxyphenyl and the trifluoropropyl group with an *iso*-butyl group gave our most potent compound 5, which had a half-maximum effective concentration (EC₅₀) value of 0.014 μM in the intra-macrophage assay. Compound 5 was metabolically unstable, although it demonstrated more than 95% parasite reduction when dosed in a hepatic cytochrome P450

¹Drug Discovery Unit, Wellcome Centre for Anti-Infectives Research, Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee, UK. ²Global Health R&D, GlaxoSmithKline, Tres Cantos, Spain. ³David Jack Centre for R&D, GlaxoSmithKline, Ware, UK. ⁴Division of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster, UK. ⁵Wellcome Sanger Institute, Cambridge, UK. ⁶Cellzome GmbH, A GlaxoSmithKline Company, Heidelberg, Germany. ⁷GlaxoSmithKline, New Frontiers Science Park, Harlow, UK.

⁸Global Health R&D, GlaxoSmithKline, Stockley Park West, Uxbridge, UK. ⁹Present address: Centre of Immunobiology, Institute of Infection, Immunity & Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. *e-mail: tim.j.miles@gsk.com; k.read@dundee.ac.uk; i.h.gilbert@dundee.ac.uk

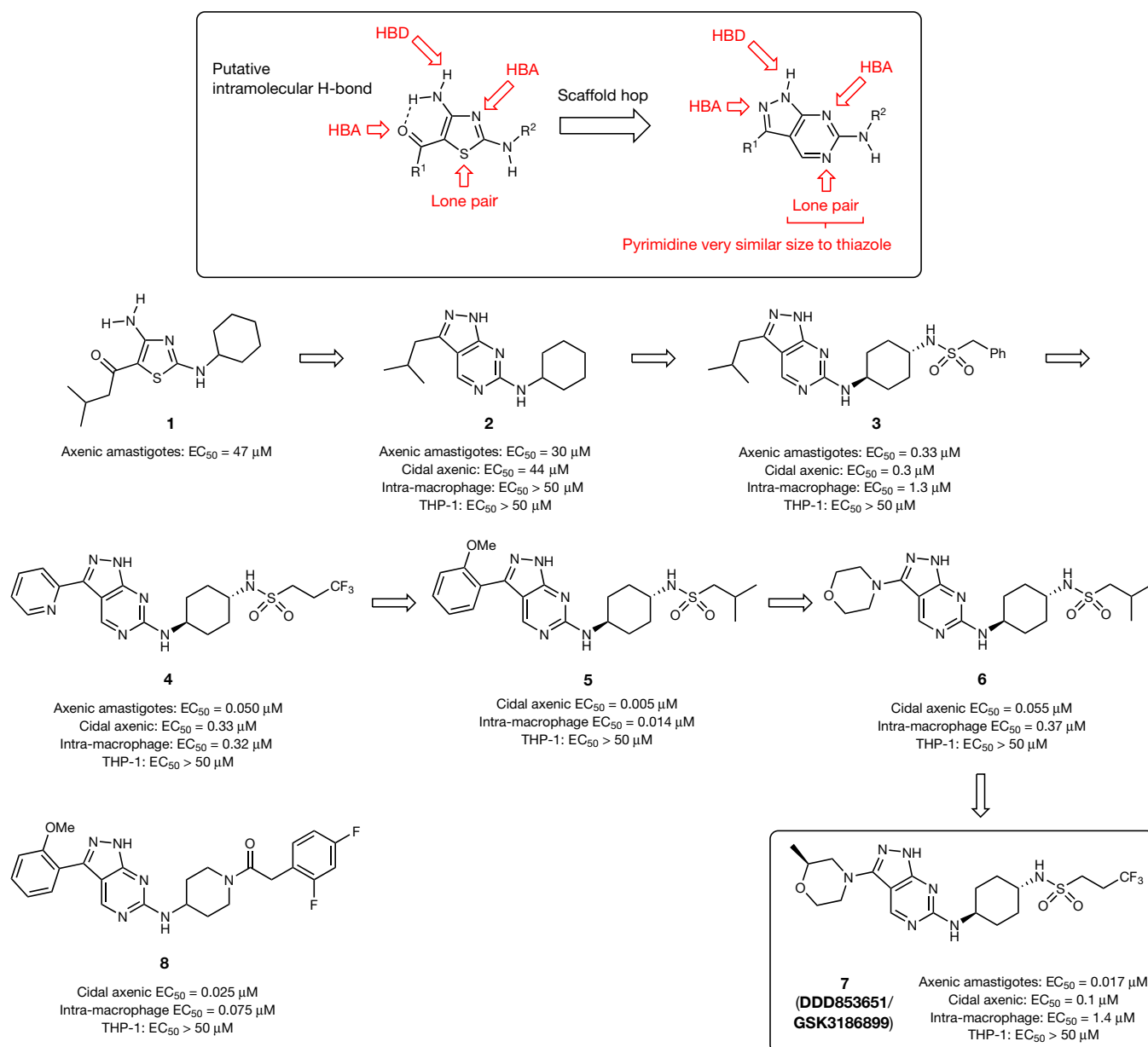


Fig. 1 | The evolution of the pyrazolopyrimidine series to give the development compound 7. Potencies against axenic amastigotes, intra-macrophage amastigotes and against THP-1 cells are shown⁹; data show the geometric mean of at least three independent replicates. In the cidal

axenic assay, a higher cell density and improved detection limit is used compared to the axenic assay, allowing distinction between cytostatic and cytotoxic compounds¹⁰. HBA, hydrogen bond acceptor; HBD, hydrogen bond donor.

reductase-null (HRN) mouse model of infection¹¹. In addition, the solubility of compounds 4 and 5 was poor.

The 2-methoxyphenyl group of 5 was replaced by a morpholine (compound 6) to increase the polarity, increase the three-dimensional shape (sp^3 character) and reduce the number of aromatic rings. This was substituted with a 2-methyl group to reduce the planarity further, and the trifluoropropyl sulfonamide was re-introduced, to give the key compound 7 (also known as **DDD853651** or **GSK3186899**)¹². This compound was selected as our preclinical candidate, on the basis of the overall properties of the molecule (potency, efficacy in the mouse model, pharmacokinetics and safety profile).

Compound 7 was active against *L. donovani* in an intra-macrophage assay⁹ with an EC_{50} value of $1.4 \mu M$ (95% confidence interval 1.2 – $1.5 \mu M$, $n = 12$), and showed good selectivity against mammalian THP-1 host cells (EC_{50} value $> 50 \mu M$). This is not as potent as our reported data for amphotericin B (EC_{50} value of $0.07 \mu M$ in the intra-macrophage assay), but is comparable to the clinically used drugs miltefosine and paromomycin (EC_{50} values of $0.9 \mu M$ and

$6.6 \mu M$, respectively)⁹. Compound 7 was also active in our cidal axenic amastigote assay (EC_{50} value of $0.1 \mu M$; 95% confidence interval 0.06 – $0.170 \mu M$, $n = 4$)¹⁰. At a concentration of $0.2 \mu M$, compound 7 was cytotoxic at 96 h; increasing the concentration to $1.8 \mu M$ reduced this time to 48 h (Extended Data Fig. 1). Compound 7 demonstrated a less than 10-fold variation in potency against a panel of *Leishmania*-derived lines. The compound was also more active in a panel of *Leishmania* lines using human peripheral blood mononuclear cells as the host cells (Extended Data Table 1).

A balance between solubility in relevant physiological media (Extended Data Table 2) and in vitro potency proved key for the development of this series. Compound 7 was stable when incubated with liver microsomes or hepatocytes, predictive of good metabolic stability (Extended Data Table 3). The compound was orally bioavailable and showed a linearity of pharmacokinetics from 10 to 300 mg kg^{-1} in rats (Extended Data Table 4). In our mouse model of infection, the compound demonstrated comparable activity to the front-line drug miltefosine, reducing parasite levels by 99% when dosed orally twice

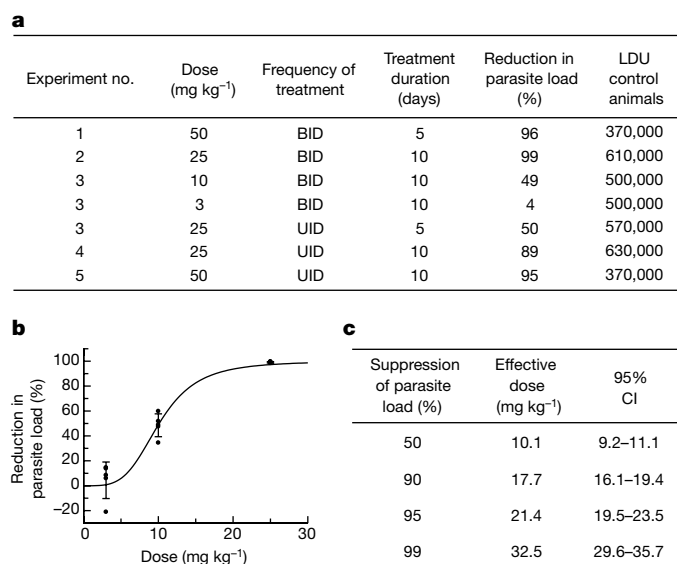


Fig. 2 | Efficacy of compound 7 in a mouse model of visceral leishmaniasis. Each arm was carried out with five mice. **a**, Reduction in parasite load for various dose regimens. BID, twice daily dosing; UID, once daily dosing. LDU, Leishman–Donovan units (the number of amastigotes per 500 nucleated cells multiplied by the organ weight in grammes^{31,32}). **b**, Dose–response curve for twice daily dosing for 10 days. Error bars show the s.d. for $n = 5$ mice. **c**, Given dose required to give a particular reduction in parasite load for twice daily dosing for 10 days. The reported ED₉₀ value for miltefosine in a mouse model is 27 mg kg⁻¹ once a day^{6,33,34}. CI, confidence interval.

a day for 10 days at 25 mg kg⁻¹ (Fig. 2). The efficacy of treatment was dependent on dose, frequency (twice a day better than once), and duration (10 days better than 5). The non-clinical safety data for compound 7 suggest a suitable therapeutic window for progression into regulatory preclinical studies. In vitro assays demonstrated that this compound does not markedly inhibit cytochrome P450 enzymes, mitigating a potential risk of problematic drug–drug interactions that is particularly relevant owing to the frequency of co-infections of visceral leishmaniasis and HIV¹.

As the series was developed from a known protein kinase scaffold¹³, Kinobead technology was used to determine whether compound 7 inhibits human protein kinases¹⁴. These experiments indicated that compound 7 interacted with four human kinases—MAPK11, NLK, MAPK14 and CDK7—at concentrations within multiples of the predicted clinical dose (Supplementary Table 1). However, the extent of inhibition of these human kinases is not sufficient to preclude clinical development of the molecule, and no marked inhibition of other human kinases was detected in the Kinobead assays. Non-GLP preclinical assessment of cardiovascular effects and genotoxicity did not reveal any issues that would prevent further development. In addition, there were no notable adverse effects in a rat seven-day repeat-dose oral toxicity study with respect to clinical chemistry and histopathology at all doses tested. Both the in vivo efficacy and safety profile of compound 7 support progression to definitive safety studies.

Mode of action studies

Determining the mode of action of new chemical series can greatly benefit drug discovery campaigns¹⁵. Because there is no blueprint to establish the mode of action of bioactive small molecules^{16,17}, several complementary methods were used. Representative pyrazolopyrimidine analogues (4, 5, 6 and 7) from the drug discovery program were used as chemical tools (Fig. 1), as was compound 8, in which the diaminocyclohexyl group was replaced by an aminopiperidine amide. These compounds showed very good activity correlation between the intra-macrophage, axenic amastigote and promastigote assays, giving us confidence to use the extracellular parasite forms (promastigote) for

mode-of-action studies where it was not possible to use the intracellular forms (amastigote) (Supplementary Table 2).

As a first step towards identifying the target(s) of the leishmanocidal pyrazolopyrimidine series, structure–activity relationships were used to inform the design of analogues containing a polyethyleneglycol (PEG) linker (9, 11 and 12; Extended Data Fig. 2), which were then covalently attached to magnetic beads to allow for chemical proteomics. First, beads derivatized with compound 9 were used to pull-down proteins from SILAC (stable isotope labelling by amino acids in cell culture)-labelled *L. donovani* promastigote lysates¹⁸ in the presence ('light-labelled lysate') or absence ('heavy-labelled lysate') of 10 μ M compound 10, a structurally related, bioactive derivative of compound 9¹⁹. After combining the bead eluates and performing proteomic analyses, proteins that bound specifically to the pyrazolopyrimidine pharmacophore could be distinguished from proteins that bound non-specifically to the beads by virtue of high heavy: light tryptic peptide isotope ratios. These experiments identified CRK12, CRK6, CYC9, CRK3, MPK9, CYC6 and a putative STE11-like protein kinase (LinJ.24.1500) as specific binders to the compound 9-derivatized beads (log₂ of heavy:light ratio > 2.8; 7-fold enrichment) (Supplementary Fig. 5 and Supplementary Table 3). Second, pull-down experiments were conducted with beads derivatized with compounds 9, 11 or 12, followed by competition studies with compounds 5, 8 or 8, respectively. Adherent proteins were washed off the beads, digested with trypsin and labelled with isobaric tandem mass tags. Comparison of the labelled peptides derived from experiments, with and without competition, by liquid chromatography–tandem mass spectrometry identified proteins that are likely to bind specifically to the immobilized ligands. Potential candidates identified included CRK3, CRK6, CRK12, CYC3, CYC6, CYC9, MPK9 and MPK5 and several hypothetical proteins (Supplementary Fig. 6 and Supplementary Table 4). We also investigated immobilizing the compound at an alternative position on the scaffold and this gave a similar binding profile (Supplementary Fig. 6 and Supplementary Table 4), further validating the approach. These results are consistent with previous studies that report that the pyrazolopyrimidine core binds to protein kinases^{13,20–22}.

The presence of cdc2-related kinases (CRK3, CRK6 and CRK12) and cyclins (CYC3, CYC6 and CYC9) in the initial target list led us to analyse the effects of pyrazolopyrimidines (5, 6, 7 and 8) on cell-cycle progression in *L. donovani*. Treatment resulted in an accumulation of cells in the G1 and in G2/M phases of the cell cycle, and a decrease in the proportion of cells in S phase (Fig. 3a for compound 7 and Supplementary Fig. 9 for compounds 5, 6 and 8), suggesting arrests in the cell-cycle at G1/S and G2/M phases, consistent with a mode of action via CRK and/or CYC components.

Resistance was generated in *L. donovani* promastigotes against compounds 4 and 5. A single cloned parental cell line was divided into three individual cultures for each compound, and resistance was generated by exposing parasites to step-wise increasing the concentrations of the compounds. After resistance generation, each independently generated cell line was cloned and three individual clones for each compound (six in total) were selected for in-depth study. The resulting clones demonstrated more than 500-fold and 9–17-fold resistance to compounds 4 and 5, respectively (Extended Data Table 5). Resistance to both compounds was found to be stable over 50 days in culture in the absence of drug pressure and, notably, all clones showed cross-resistance to compounds 4 and 5, and 20–50-fold cross-resistance to compound 7. These data suggest our pyrazolopyrimidines share common mechanisms of resistance and most likely modes of action. Importantly, intracellular amastigotes, derived from the resistant promastigotes, were 8.5-fold and 5-fold resistant to compounds 5 and 7, respectively, compared to wild-type parasites (Extended Data Table 6), strongly suggesting that their mechanism(s) of action are the same in promastigote and intracellular amastigote stages of the parasite.

To gain further insight into the mechanism of action and potential target(s) of this pyrazolopyrimidine series, our six drug-resistant clones underwent whole-genome sequence analysis. A range of mutations, relative to parental clones, were found across the genome (Supplementary Table 5), including a long region with loss of heterozygosity on

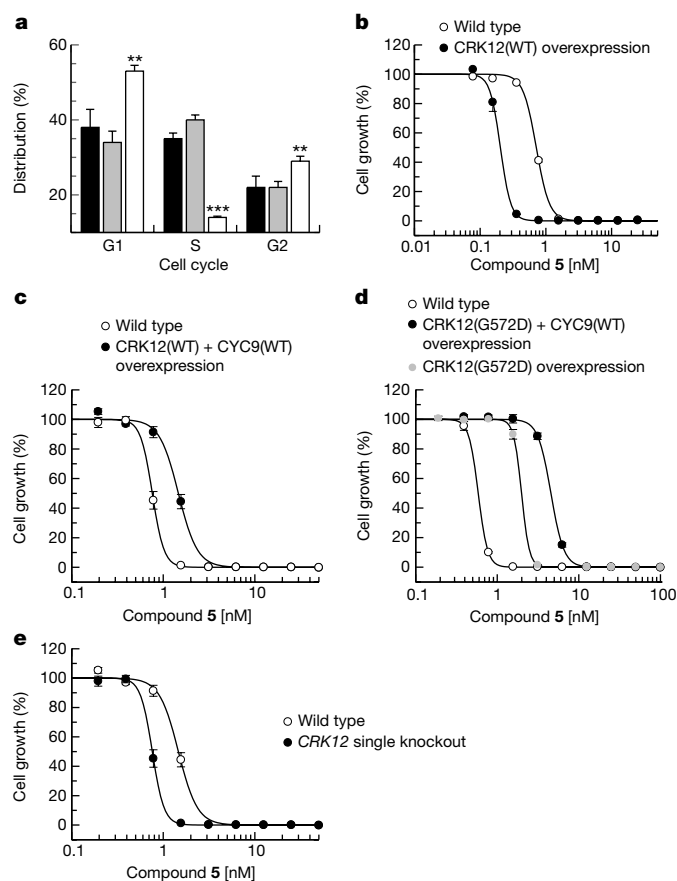


Fig. 3 | Studies to validate the molecular target of the pyrazolopyrimidine series. **a**, Cell cycle analysis after treatment with compound 7 for 8 h. Untreated cells at 0 h (black) and at 8 h (grey). Cells were treated with 5 × the EC₅₀ value of compound 7 for 8 h (white). ***P* = 0.01; ****P* = 0.001 (unpaired Student's *t*-test). **b**, Effects of overexpression of wild-type CRK12 in promastigotes on the potency of compound 5 (EC₅₀ value of 0.24 ± 0.002 nM (mean ± s.d.), closed circles) compared to wild-type cells treated with compound 5 (0.72 ± 0.01 nM, open circles). **c**, Effects of co-overexpression of wild-type CRK12 and CYC9 in promastigotes on the potency of compound 5 (EC₅₀ value of 1.43 ± 0.01 nM, closed circles) compared to wild-type cells treated with compound 5 (EC₅₀ value of 0.5 ± 0.004 nM, open circles). **d**, Effects of overexpression of mutant CRK12(G572D) and CYC9 in promastigotes on the potency of compound 5 (EC₅₀ value of 4.6 ± 0.05 nM, closed circles) compared to wild-type cells treated with compound 5 (EC₅₀ value of 0.59 ± 0.001 nM, open circles) and promastigotes overexpressing CRK12(G572D) alone (EC₅₀ value of 1.99 ± 0.002 nM, grey circles). **e**, Effect of knocking out a single copy of the *CRK12* gene on the potency of compound 5 in promastigotes (EC₅₀ value of 0.76 ± 0.004 nM, closed circles) compared to wild-type cells treated with compound 5 (EC₅₀ value of 1.5 ± 0.004 nM, open circles). *P* = 0.0014 (wild type compared to *CRK12* knockout), unpaired Student's *t*-test. Data are mean ± s.d. from three technical replicates and are representative of at least duplicate experiments.

chromosome 9. In total, 75 sites were identified genome-wide that each had single base substitutions that resulted in a non-synonymous change in at least one clone (Supplementary Table 6). Most (65) of the non-synonymous substitutions consisted of derived clones losing a parental allele but amplifying the remaining allele. In five of the six resistant clones, a new heterozygous substitution was selected in a single gene of unknown function (*LdBPK_251630*) but most notably, in all six drug-resistant clones, a single homozygous non-synonymous substitution was found in *CRK12* (*LdBPK_090270*), a gene within the long loss-of-heterozygosity region. This mutation changes Gly572 to Asp and falls within the region predicted to encode the catalytic domain of *L. donovani* CRK12. This suggests that CRK12 is the target of the pyrazolopyrimidines. Extensive variations in chromosomal copy

numbers are common in *Leishmania*^{23,24}, and extra copies of chromosome 9, containing the *CRK12* gene, were found in four out of six of the drug-resistant clones (Supplementary Table 7). In addition, three of these four clones had extra copies of chromosome 32, containing the gene for *CYC9*. Previous studies in *T. brucei* have established that the partner cyclin of CRK12 is *CYC9*²⁵. This suggests that *CYC9* may be the cognate cyclin partner for *L. donovani* CRK12.

Target validation

To dissect the role of CRK12 and *CYC9* in the mechanism of action and resistance of pyrazolopyrimidines, a series of protein overexpression studies were undertaken in *L. donovani* promastigotes. In all cases, overexpression of putative targets was confirmed by increased levels of transcripts in our transgenic cell lines relative to the wild-type cells, as determined by quantitative PCR (qPCR) (Supplementary Table 8).

Counterintuitively, overexpression of wild-type CRK12 rendered the parasites approximately 3-fold more sensitive to compound 5 (Fig. 3b). The overexpression of *CYC9* alone had no effect on compound resistance, but co-overexpression of *CYC9* and wild-type CRK12 rendered the transgenic parasites around 3-fold more resistant to compounds 5 and 7 (Fig. 3c and Supplementary Table 8). Next, we looked at the mutated (Gly572 to Asp) version of CRK12 (CRK12(G572D)) identified in all of our drug-resistant clones. Overexpression of CRK12(G572D) rendered the parasites around 3.4-fold resistant to compound 5 (Fig. 3d and Supplementary Table 8) and to the preclinical lead compound 7 (Supplementary Table 8), while being equally sensitive to the unrelated nitroimidazole drug fexinidazole sulfone (Supplementary Table 9). Co-overexpression of CRK12(G572D) and *CYC9* rendered the parasites 6-fold more resistant to compound 7 and 8-fold more resistant to compound 5. This shift in sensitivity is considerably greater than the 3.4-fold increased resistance observed with parasites overexpressing CRK12(G572D) alone (Fig. 3d). Replacing a single copy of the *CRK12* gene with a drug-selectable marker left parasites approximately 2-fold more susceptible to compound 5 than the wild-type cells (Fig. 3e and Supplementary Fig. 10). We were unable to directly replace both endogenous copies of the *CRK12* gene, except in the presence of an ectopic copy of the gene, suggesting that the *CRK12* gene is essential for the growth and survival of *L. donovani* (Supplementary Fig. 10).

Initially, CRK3 and CRK6 were identified as credible targets based on our collective proteomics datasets, as well as their established roles in kinetoplast cell cycle regulation^{26,27}. However, whole-genome sequencing, qPCR (Supplementary Fig. 8) and Southern blot (Supplementary Fig. 7) analysis of resistant clones confirmed that mutations within, or amplification of, the *CRK3* and *CRK6* genes were not responsible for resistance to pyrazolopyrimidines. Direct modulation of CRK3 and CRK6 levels within *L. donovani* promastigotes, by generating overexpressing and single-gene knockout parasites, did not alter drug sensitivity (Supplementary Table 8). Overexpression of CRK3 and CRK6 in combination with their cognate cyclin partners *CYC6* and *CYC3* was not possible because co-overexpression proved toxic. Collectively, these data suggest that the primary mechanism of action of this compound series is unlikely to be via CRK3 or CRK6 inhibition.

Commonly, overexpression of the molecular target of a compound is accompanied by an increase in drug resistance. With this in mind, our collective data strongly suggest that the principal target of our pyrazolopyrimidine series is the *CYC9*-activated form of CRK12, such that overexpression of CRK12 and *CYC9* together provides resistance. This hypothesis is also consistent with the amplification of both *CRK12* and *CYC9* in resistant parasites; as well as the identification of both proteins in our SILAC and Kinobead proteomic datasets. The fact that overexpression of *CYC9* alone has no effect suggests that *CYC9* is, to some extent, in excess over CRK12 and thus overexpression of CRK12(G572D) can provide (3-fold) resistance that is increased when additional *CYC9* is co-expressed (8-fold). The 'hyper-sensitivity' of parasites overexpressing wild-type CRK12 alone to these compounds remains perplexing. One potential explanation is that wild-type CRK12 bound to a

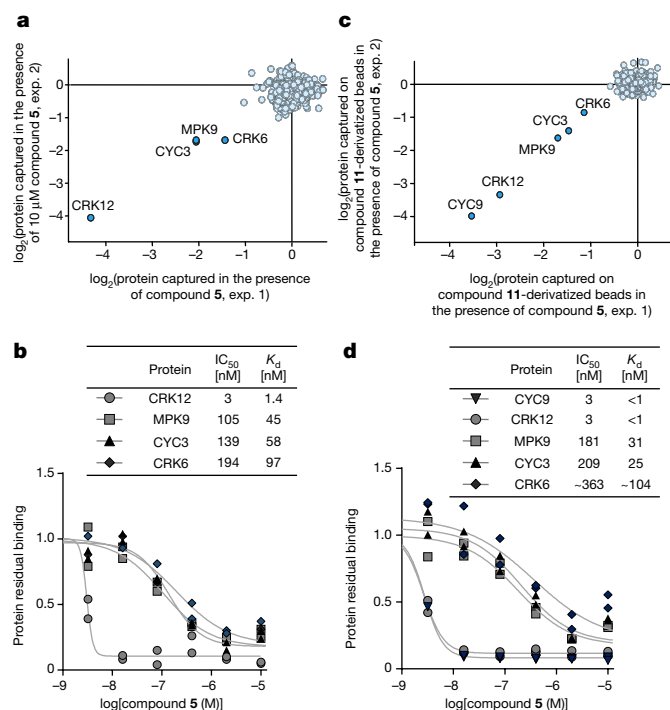


Fig. 4 | Identification of cyclin-dependent related kinases as targets of the pyrazolopyrimidine series using a chemoproteomic approach.

a, Relative amounts of protein captured on Kinobeads in the presence of 10 μ M compound **5** compared to vehicle; data are representative of two experiments (experiment 1 on the x axis, experiment 2 on the y axis). A log₂ scale is used. **b**, Dose–response curves of proteins binding to Kinobeads in the presence of varying concentrations of compound **5**. Apparent K_d values are shown. **c**, Relative amounts of protein captured on 11-derivatized beads in the presence of 10 μ M compound **5** compared to vehicle; data are representative of two experiments (experiment 1 on the x axis, experiment 2 on the y axis). A log₂ scale is used. **d**, Dose–response curves of binding of proteins to 11-derivatized beads in the presence of varying concentrations of compound **5**.

pyrazolopyrimidine in the absence of a CYC9 subunit is particularly toxic to the parasite. Alternatively, increased levels of CRK12 may well sequester other cyclins, thereby preventing their essential interactions with other CRKs. Further studies will be required to test these hypotheses.

Given that the compounds from this chemical series interacted with protein kinases, in particular CRK12, we used Kinobead technology^{14,28,29} with axenic amastigote extracts to identify pyrazolopyrimidine-binders in the *Leishmania* kinome. These experiments were performed in the presence or absence of an excess of the soluble parent compound **5**. All proteins captured by the beads were quantified by tandem mass tag (TMT) labelling of tryptic peptides followed by liquid chromatography–tandem mass spectrometry analysis³⁰. CRK12, MPK9, CRK6 and CYC3 (Fig. 4a) were identified, consistent with the other experiments above. A dose–response experiment was performed in which compound **5** was added over a range of concentrations to establish a competition-binding curve and determine a half-maximal inhibition concentration (IC₅₀) value (Fig. 4b). The IC₅₀ values obtained in these experiments represent a measure of target affinity, but are also affected by the affinity of the target for the bead-immobilized ligands. The latter effect can be deduced by determining the depletion of the target proteins by the beads, such that apparent dissociation constant (K_d) values can be determined that are largely independent from the bead ligand³⁰. The apparent K_d values were 1.4 nM for CRK12, 45 nM for MPK9, 58 nM for CYC3 and 97 nM for CRK6. These values are determined in physiological conditions (substrates, cyclins and ATP) and provide further compelling evidence that the principal target of this compound series is CRK12. Further pull-downs with a resin-bound pyrazolopyrimidine analogue (**11**) were conducted in parallel with the Kinobead experiments and returned broadly similar results (Fig. 4c, d).

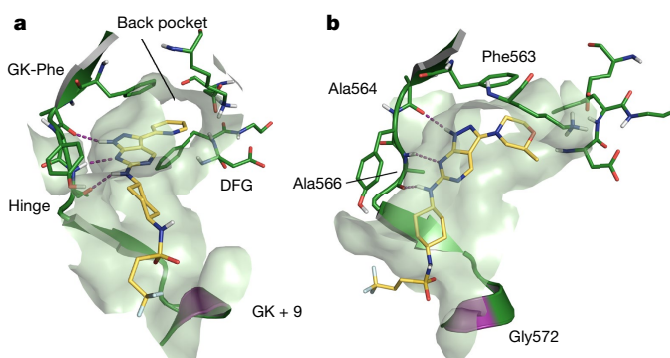


Fig. 5 | Docking studies for compounds **4 and **7**.** **a**, **b**, Docking poses for compound **4** (**a**) and compound **7** (**b**). Dotted purple lines represent hydrogen bonds. The mutated residue in position gatekeeper (GK) +9 is indicated in purple in the ribbon diagram. DFG denotes the conserved Asp-Phe-Gly motif.

Collectively, our data provide strong evidence that CRK12 forms an important interaction with CYC9: (1) our studies indicate that over-expression of CYC9 together with CRK12 markedly increases resistance to our pyrazolopyrimidine compounds; (2) in several of our compound-resistant cell lines, additional copies of chromosome 32, containing the CYC9 gene, were found; (3) in the related organism *T. brucei*, CYC9 was confirmed as the partner cyclin for CRK12; and (4) in several chemical proteomics studies, CYC9 was identified as binding to immobilized compounds from our pyrazolopyrimidines alongside CRK12.

Modelling

A homology model was built for *L. donovani* CRK12 using the structure of human cyclin dependent kinase 9 (CDK9, Protein Data Bank (PDB) code 4BCF) as a template. (Notably, compound **7** showed an IC₅₀ value of greater than 20 μ M against CDK9 in the Kinobead assay.) A combination of docking studies, molecular dynamics simulation and free-energy calculations indicated that the most likely binding mode is that shown in Fig. 5 (see Supplementary Information for discussion). With very few exceptions, the binding modes of protein kinase inhibitors are highly conserved across kinase family members; searching the protein database revealed a related 5-amino-pyrazolopyrimidine, which bound to ALK in a very similar fashion (PDB code 4Z55, ligand 4LO). In our proposed binding mode, the bicyclic scaffold interacts with the hinge residues establishing two hydrogen bonds between the *sp*² pyrimidine nitrogen in position 6 and the backbone NH of Ala566, and between the pyrazole NH in position 1 and the backbone carbonyl oxygen of Ala564 (Fig. 5b). A third hydrogen bond is also established between the amino NH in position 5 and the backbone carbonyl oxygen of Ala566. The substituent in position 3 of the pyrazole ring is directed towards the ATP back-pocket interfacing with the gatekeeper residue (Phe563). This binding mode is consistent with the analogues **9**, **11** and **12** retaining binding affinity, with the PEG linkers being attached to water-accessible parts of the core. The Gly572Asp mutation that causes resistance to the pyrazolopyrimidine series is located at the end of the hinge region, nine residues from the gatekeeper. In the Gly572Asp mutant, the negatively charged side chain of the aspartic acid is positioned in close contact to the oxygen atoms of the sulfonamide moiety, leading to an unfavourable electrostatic interaction.

Discussion

New oral drugs for visceral leishmaniasis, particularly those capable of treating ongoing outbreaks in East Africa, are urgently needed. Effective drugs will make a notable difference to treatment outcomes for this devastating parasitic disease. With the ultimate goal of elimination of visceral leishmaniasis, several treatment options will be required. We have identified a pyrazolopyrimidine series that shows the potential to treat visceral leishmaniasis. Our studies indicate that the principal mechanism of action of our pyrazolopyrimidine compounds is by the inhibition of CRK12, defining CRK12 as one of very few chemically

validated drug targets in *Leishmania*. Furthermore, our data indicate that CYC9 is the definitive partner cyclin for CRK12. The physiological function(s) of CRK12 and CYC9 have yet to be determined and the availability of our inhibitory pyrazolopyrimidines should assist in probing this aspect of parasite biology.

It is clear from our collective chemical proteomics studies that the pyrazolopyrimidines also interact with other *Leishmania* protein kinases, in particular CRK6 and CRK3, albeit with considerably lower affinities than for CRK12. Although CRK12 is undoubtedly the principal target of this compound series, we cannot rule out the possibility that underlying this mechanism of action is an element of polypharmacology. Indeed, the inhibition of secondary kinase targets may be responsible for some of the phenotypic effects observed in drug-treated parasites, such as cell cycle arrest.

Compound 7 is being advanced towards human clinical trials and is currently undergoing preclinical development. The data generated so far provide a reason to believe that compound 7 has the potential to fulfil the community target product profile (see <https://www.dndi.org/diseases-projects/leishmaniasis/tpp-v1>). However, as a systematic approach to drug discovery is relatively new in this neglected disease, and there is a lack of correlation between preclinical and clinical data, there are outstanding questions that can only be answered as the compound progresses through development.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Compound 7 is currently in preclinical development and full disclosure of the synthesis of this compound has been included in this publication. All reasonable requests for the other key tool molecules disclosed in this manuscript will be met subject to an appropriate material transfer agreement in place between all parties.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0356-z>.

Received: 9 August 2017; Accepted: 11 June 2018;

Published online 25 July 2018.

- Alvar, J. et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* **7**, e35671 (2012).
- Ritmeijer, K. & Davidson, R. N. Royal Society of Tropical Medicine and Hygiene joint meeting with Médecins Sans Frontières at Manson House, London, 20 March 2003: field research in humanitarian medical programmes. Médecins Sans Frontières interventions against kala-azar in the Sudan, 1989–2003. *Trans. R. Soc. Trop. Med. Hyg.* **97**, 609–613 (2003).
- Sundar, S. et al. Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use. *Clin. Infect. Dis.* **55**, 543–550 (2012).
- den Boer, M. L., Alvar, J., Davidson, R. N., Ritmeijer, K. & Balasegaram, M. Developments in the treatment of visceral leishmaniasis. *Expert Opin. Emerg. Drugs* **14**, 395–410 (2009).
- Mueller, M. et al. Unresponsiveness to AmBisome in some Sudanese patients with kala-azar. *Trans. R. Soc. Trop. Med. Hyg.* **101**, 19–24 (2007).
- Khare, S. et al. Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature* **537**, 229–233 (2016).
- Don, R. & Ioset, J.-R. Screening strategies to identify new chemical diversity for drug development to treat kinetoplastid infections. *Parasitology* **141**, 140–146 (2014).
- Woodland, A. et al. From on-target to off-target activity: identification and optimisation of *Trypanosoma brucei* GSK3 inhibitors and their characterisation as anti-*Trypanosoma brucei* drug discovery lead molecules. *ChemMedChem* **8**, 1127–1137 (2013).
- De Rycker, M. et al. Comparison of a high-throughput high-content intracellular *Leishmania donovani* assay with an axenic amastigote assay. *Antimicrob. Agents Chemother.* **57**, 2913–2922 (2013).
- Nühs, A. et al. Development and validation of a novel *Leishmania donovani* screening cascade for high-throughput screening using a novel axenic assay with high predictivity of leishmanicidal intracellular activity. *PLoS Negl. Trop. Dis.* **9**, e0004094 (2015).
- Henderson, C. J., Pass, G. J. & Wolf, C. R. The hepatic cytochrome P450 reductase null mouse as a tool to identify a successful candidate entity. *Toxicol. Lett.* **162**, 111–117 (2006).
- Miles, T. J. & Thomas, M. G. Pyrazolo[3,4-d]pyrimidin derivative and its use for the treatment of leishmaniasis. WIPO patent WO/2016/116563 (2016).
- Ding, Q., Jiang, N. & Roberts, J. L. Pyrazolo pyrimidines. WIPO patent WO/2005/121107 (2005).
- Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–1044 (2007).
- Terstappen, G. C., Schlüpen, C., Raggiaschi, R. & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* **6**, 891–903 (2007).
- Park, J., Koh, M. & Park, S. B. From noncovalent to covalent bonds: a paradigm shift in target protein identification. *Mol. Biosyst.* **9**, 544–550 (2013).
- Lee, H. & Lee, J. W. Target identification for biologically active small molecules using chemical biology approaches. *Arch. Pharm. Res.* **39**, 1193–1201 (2016).
- Ursu, A. & Waldmann, H. Hide and seek: identification and confirmation of small molecule protein targets. *Bioorg. Med. Chem. Lett.* **25**, 3079–3086 (2015).
- Urbanianak, M. D., Guther, M. L. S. & Ferguson, M. A. J. Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages. *PLoS ONE* **7**, e36619 (2012).
- Liu, Y. & Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2**, 358–364 (2006).
- Zhang, L. et al. Design, synthesis, and biological evaluation of pyrazolopyrimidine-sulfonamides as potent multiple-mitotic kinase (MMK) inhibitors (part I). *Bioorg. Med. Chem. Lett.* **21**, 5633–5637 (2011).
- Freyne, E. J. E. et al. Pyrazolopyrimidines as cell cycle kinase inhibitors. WIPO patent WO/2006/074984 (2006).
- Rogers, M. B. et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* **21**, 2129–2142 (2011).
- Downing, T. et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **21**, 2143–2156 (2011).
- Monnerat, S. et al. Identification and functional characterisation of CRK12:CYC9, a novel cyclin-dependent kinase (CDK)–cyclin complex in *Trypanosoma brucei*. *PLoS ONE* **8**, e67327 (2013).
- Hassan, P., Fergusson, D., Grant, K. M. & Mottram, J. C. The CRK3 protein kinase is essential for cell cycle progression of *Leishmania mexicana*. *Mol. Biochem. Parasitol.* **113**, 189–198 (2001).
- Tu, X. & Wang, C. C. Pairwise knockdowns of cdc2-related kinases (CRKs) in *Trypanosoma brucei* identified the CRKs for G1/S and G2/M transitions and demonstrated distinctive cytokinetic regulations between two developmental stages of the organism. *Eukaryot. Cell* **4**, 755–764 (2005).
- Médard, G. et al. Optimized chemical proteomics assay for kinase inhibitor profiling. *J. Proteome Res.* **14**, 1574–1586 (2015).
- Bergamini, G. et al. A selective inhibitor reveals PI3K γ dependence of T_H17 cell differentiation. *Nat. Chem. Biol.* **8**, 576–582 (2012).
- Bantscheff, M. et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255–265 (2011).
- Bradley, D. J. & Kirkley, J. Regulation of *Leishmania* populations within the host. I. The variable course of *Leishmania donovani* infections in mice. *Clin. Exp. Immunol.* **30**, 119–129 (1977).
- Croft, S. L., Snowdon, D. & Yardley, V. The activities of four anticancer alkyllylphospholipids against *Leishmania donovani*, *Trypanosoma cruzi* and *Trypanosoma brucei*. *J. Antimicrob. Chemother.* **38**, 1041–1047 (1996).
- Seifert, K. & Croft, S. L. In vitro and in vivo interactions between miltefosine and other antileishmanial drugs. *Antimicrob. Agents Chemother.* **50**, 73–79 (2006).
- Escobar, P., Yardley, V. & Croft, S. L. Activities of hexadecylphosphocholine (miltefosine), AmBisome, and sodium stibogluconate (Pentostam) against *Leishmania donovani* in immunodeficient scid mice. *Antimicrob. Agents Chemother.* **45**, 1872–1875 (2001).

Acknowledgements The authors acknowledge the Wellcome Trust for funding (grants 092340, 105021, 100476, 101842, 079838 and 098051).

Reviewer information Nature thanks R. Guy, J. Mottram and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions In brief, S.W., M.D.U., T.D.O., H.P., M.Bo. and S.M. carried out the mode of action, genomic and proteomic studies. M.T., S.P. and S.A. carried out the chemistry studies. M.D.R., S.M., L.M.M. and L.Sa. carried out the parasite screening. S.C., L.S., F.R.C.S. and P.C. carried out the drug metabolism and pharmacokinetic studies. F.Z. and N.H. carried out the molecular modelling. R.L. and S.G. carried out the safety studies. S.W., M.T., S.P., M.D.R., R.L., S.G., M.D.U., L.M.M., F.Z., M.Be., G.D., D.W.G., S.G.-D., S.D., J.M.F., P.W.G., M.A.J.F., A.H.F., T.J.M., K.D.R. and I.H.G. designed experiments, managed parts of the project and contributed to the writing. See Supplementary Information for further details.

Competing interests These authors have shares in GlaxoSmithKline: P.G.W., S.D., T.J.M., K.D.R., S.C., R.L., S.G., M.Bo., H.P., P.C., G.D., D.G., S.G.-D. and J.M.F. The other authors declare no competing interests.

Additional information

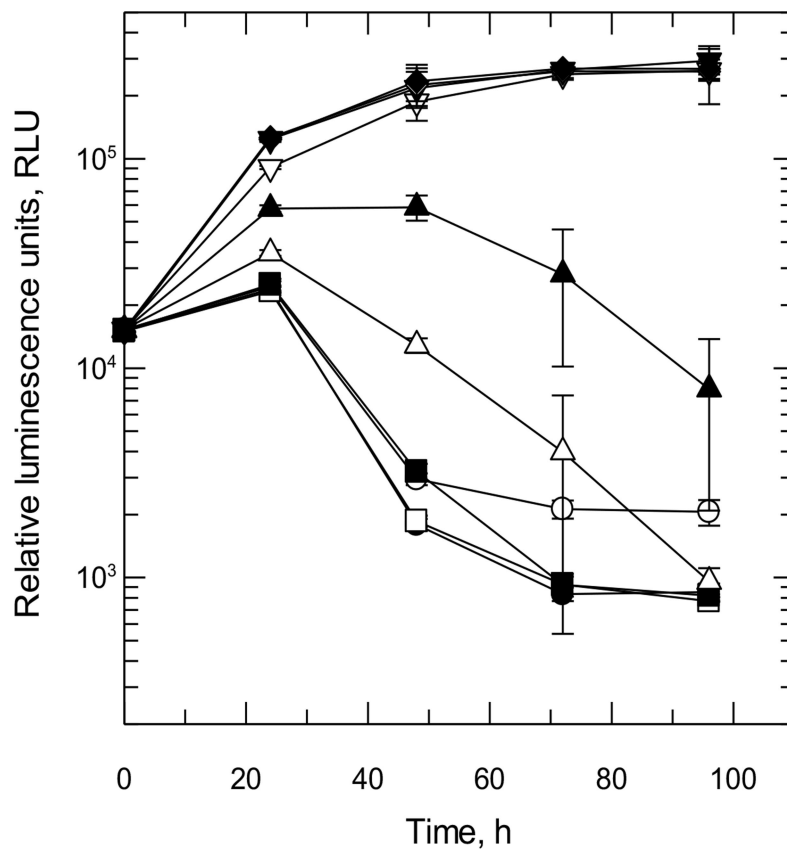
Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0356-z>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0356-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

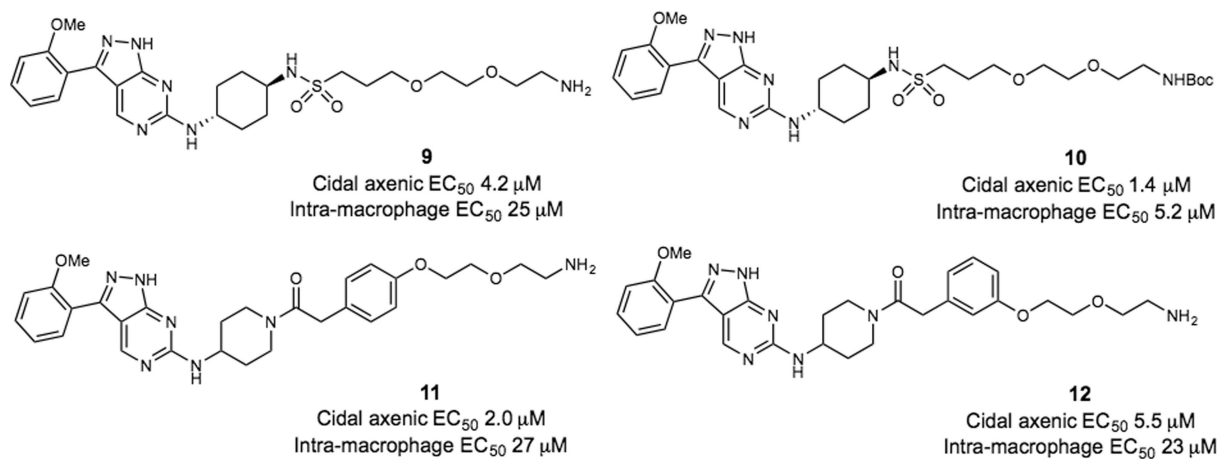
Correspondence and requests for materials should be addressed to T.J.M., K.D.R. or I.H.G.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Rate-of-kill of *L. donovani* axenic amastigotes by compound 7. Chart shows relative luminescence units (RLU) versus time from axenic amastigote rate-of-kill experiment with compound 7 (representative results for one of two independent experiments are shown; data are mean and s.d. of three technical replicates). Concentrations are as

follows (μM): 50, open circles; 16.7, closed circles; 5.6, open squares; 1.85, closed squares; 0.62, open triangles; 0.21, closed triangles; 0.069, open inverted triangles; 0.023, closed inverted triangles, 0.0076, open diamonds and 0.0025, closed diamonds.



Extended Data Fig. 2 | Linker-containing target molecules synthesized for chemical proteomic experiments and their corresponding EC_{50} values. Potencies of the compounds in the cidal axenic and

intra-macrophage assays are shown; data are from at least three independent replicates.

Extended Data Table 1 | Activity of compound 7 and miltefosine against a panel of *Leishmania* clinical isolates

Strain	Country of origin	Year	Compound 7 EC ₅₀ (μM)	Miltefosine EC ₅₀ (μM)
<i>L. donovani</i> LV9	Ethiopia	1967	0.05, 0.09	0.36, 0.43
<i>L. donovani</i> SUKA 001	Sudan	2010	0.09	0.91
<i>L. donovani</i> BHU1 *	India	2002	0.11	0.50
<i>L. donovani</i> DD8	India	1980	0.13	0.51
<i>L. infantum</i> ITMAP263	Morocco	1967	0.50, 0.13	1.0, 0.79

Intra-macrophage assays using human peripheral blood mononuclear cells are shown. Strains were tested as technical duplicates on a single (DD8, SUKA 001, BHU1) or on two (LV9, ITMAP263) occasion(s); the respective EC₅₀ values are shown for LV9 and ITMAP263. References for the cell lines are given in the Supplementary Information.

*Antimony-resistance reference strain.

Extended Data Table 2 | Solubility of compound 7 in simulated physiological media

Media	Final pH	Solubility [mg/mL]
SGF pH1.6	1.5	1.12
Fasted SIF pH6.5	6.5	0.017
Fed SIF pH6.5	6.5	0.025

SGF, simulated gastric fluid; SIF, simulated intestinal fluid. Data were generated using crystalline polymorph form 1. Solubility experiments were performed at 37 °C for 4 h.

Extended Data Table 3 | In vitro metabolic stability data for compound 7

Species	Concentration (μ M)	Liver Microsomes Cli (mL/min/g tissue)	Hepatocytes Cli (mL/min/g tissue)
Mouse	0.5	0.52	0.84
Rat	0.5	<0.5	0.77
Dog	0.5	<0.4	0.31
Human	0.5	0.71	0.55

Cli, intrinsic clearance.

Extended Data Table 4 | Drug metabolism and pharmacokinetics data for compound 7

Intravenous	Mouse (male, CD1)	Rat (male, SD)
	1 mg/kg	1 mg/kg
Cl (ml/min/kg)	169 ± 50	14 ± 9
V _{dss} (L/kg)	4.0 ± 0.5	0.4 ± 0.1
T _{1/2} (h)	0.3 ± 0.04	0.4 ± 0.2
AUC _(0-inf) (ng.h/mL)	104 ± 26	1514 ± 782
Oral	10 mg/kg	10 mg/kg
C _{max} (ng/mL)	561 ± 148	1043 ± 261
T _{max} (h)	2	2
AUC _(0-inf) (ng.h/mL)	1463 ± 362	6475 ± 2494
F% based on AUC _(0-inf)	>100	46 ± 18
Oral	100 mg/kg	100 mg/kg
C _{max} (ng/mL)	8813 ± 1966	8470 ± 3750
T _{max} (h)	3	7.3
AUC _(0-inf) (ng.h/mL)	39433 ± 23830	61202 ± 23591
F% based on AUC _(0-inf)	>100	40 ± 15
Oral	300 mg/kg	300 mg/kg
C _{max} (ng/mL)	11393 ± 4212	14833 ± 2676
T _{max} (h)	5	7.3
AUC _(0-inf) (ng.h/mL)	*66150 ± 636	136333 ± 24846
F% based on AUC _(0-inf)	>100	51 ± 22

AUC, area under the curve; Cl, clearance; C_{max}, maximum concentration; F%, oral bioavailability; V_{dss}, volume of distribution at steady-state; T_{1/2}, half-life; T_{max}, time at which maximum concentration is reached. CD1 mice and Sprague Dawley (SD) rats were used. Each arm was carried out with three animals.

*Back-extrapolated AUC greater than 20%.

Extended Data Table 5 | Sensitivity of wild-type and drug-resistant promastigotes to compounds within the series

Cell line	Compound 4		Compound 5		Compound 7	
	pEC ₅₀ (s.d.)	Fold	pEC ₅₀ (s.d.)	Fold	pEC ₅₀ (s.d.)	Fold
Wild type (Start clone)	7 (0.1)	1	8.2 (0.4)	1	7.1 (0.3)	1
Wild type (Age-matched)	7.1 (0.2)	1	8.2 (0.1)	1	7.3 (0.2)	1
4-resistant clone 1	< 4.3	>500	7.2 (0.1)	11	5.8 (0.4)	20
4-resistant clone 2	< 4.3	>500	7.3 (0.1)	7	5.7 (0.2)	24
4-resistant clone 3	< 4.3	>500	7 (0.2)	17	5.4 (0.1)	48
5-resistant clone 1	< 4.3	>500	7.1 (0.2)	11	5.5 (0.2)	41
5-resistant clone 2	< 4.3	>500	7.1 (0.2)	14	5.5 (0.1)	35
5-resistant clone 3	< 4.3	>500	7.3 (0.1)	9	5.7 (0.1)	22

Resistance was generated against compounds **4** and **5**. Values in parentheses denote s.d., $n = 3$ independent replicates.
pEC₅₀, negative logarithm of the EC₅₀ value. Potencies are mean pEC₅₀.

Extended Data Table 6 | Sensitivity of wild-type and compound 5-resistant intra-macrophage amastigotes to the compound series

Compound	Cell line	pEC₅₀	Host cell pEC₅₀	Fold difference
5	WT	7.5	<5.3	-
5	5 RES clone 1	6.6	<5.3	8.5
7	WT	5.9	<4.3	-
7	5 RES clone 1	5.2	<4.3	5.0

RES, resistant; pEC₅₀, negative logarithm of the EC₅₀ value (the compound concentration showing 50% inhibition of the growth of the cells).

New mitochondrial DNA synthesis enables NLRP3 inflammasome activation

Zhenyu Zhong^{1,2,11}, Shuang Liang^{3,4,11}, Elsa Sanchez-Lopez^{1,2}, Feng He^{1,2}, Shabnam Shalapour^{1,2}, Xue-jia Lin^{1,2,5}, Jerry Wong^{1,2}, Siyuan Ding^{6,7,8}, Ekihiro Seki⁹, Bernd Schnabl³, Andrea L. Hevener¹⁰, Harry B. Greenberg^{6,7,8}, Tatiana Kisseleva⁴ & Michael Karin^{1,2*}

Dysregulated NLRP3 inflammasome activity results in uncontrolled inflammation, which underlies many chronic diseases. Although mitochondrial damage is needed for the assembly and activation of the NLRP3 inflammasome, it is unclear how macrophages are able to respond to structurally diverse inflammasome-activating stimuli. Here we show that the synthesis of mitochondrial DNA (mtDNA), induced after the engagement of Toll-like receptors, is crucial for NLRP3 signalling. Toll-like receptors signal via the MyD88 and TRIF adaptors to trigger IRF1-dependent transcription of CMPK2, a rate-limiting enzyme that supplies deoxyribonucleotides for mtDNA synthesis. CMPK2-dependent mtDNA synthesis is necessary for the production of oxidized mtDNA fragments after exposure to NLRP3 activators. Cytosolic oxidized mtDNA associates with the NLRP3 inflammasome complex and is required for its activation. The dependence on CMPK2 catalytic activity provides opportunities for more effective control of NLRP3 inflammasome-associated diseases.

Inflammation is initiated by the sensing of pathogen-associated or damage-associated molecular patterns^{1,2}. Among pattern-recognition receptors, NLRP3 is unique in its response to highly diverse extracellular stimuli, several of which link tissue damage to sterile inflammation, the goal of which is damage repair^{3,4}. After stimulation, NLRP3 is thought to expose its pyrin domain, which binds ASC (apoptosis-associated speck-like protein) that recruits the effector molecule pro-caspase-1 via CARD–CARD interactions to form a large cytosolic complex—the NLRP3 inflammasome^{1,2,5}. Inflammasome assembly triggers the self-cleavage and activation of caspase-1, converting pro-IL-1 β and pro-IL-18 to their mature forms¹. Persistent and aberrant NLRP3 signalling underlies many chronic and degenerative diseases, including periodic auto-inflammatory syndromes, gout, osteoarthritis, Alzheimer's disease, type 2 diabetes, atherosclerosis, lupus, macular degeneration and cancer^{6,7}. To our knowledge, there are currently no effective, safe and selective therapeutic approaches to these diseases that allow the inhibition of NLRP3 inflammasome activity.

NLRP3 inflammasome activation depends on two functionally distinct steps: 'priming' and 'activation'^{1,4,8}. Priming involves the direct engagement of Toll-like receptors (TLRs) by pathogen-associated or damage-associated molecular patterns, resulting in the rapid activation of NF- κ B, which stimulates pro-IL-1 β synthesis and increased expression of NLRP3. The activation step is less clear, leading to NLRP3 inflammasome assembly followed by caspase-1 activation^{1,5,8}. A major difficulty in understanding the activation step is the abrupt transition from priming to activation that is triggered by chemically and structurally diverse stimuli (for example, microparticles, pore-forming toxins, ATP and certain pathogens), often referred to as NLRP3 activators, although none of them binds to NLRP3 directly^{1,9}. One solution to the activation conundrum is the proposal that NLRP3 activators operate through a common intracellular intermediate, most likely the

mitochondrion^{10–12}. Through different mechanisms that may involve plasma membrane rupture, K⁺ efflux and increased intracellular Ca²⁺, NLRP3 activators elicit a particular form of mitochondrial damage that causes the release of fragmented mtDNA and the increased production of reactive oxygen species (ROS) that convert mtDNA to an oxidized form (ox-mtDNA), proposed to serve as the ultimate NLRP3 ligand, or at least a part of it¹³. Mitochondrial damage induced by NLRP3 activators is distinct from that induced by pro-apoptotic BCL2 proteins, which enable the release of cytochrome *c* and activation of the apoptotic protease activating factor complex to activate caspase-3, rather than caspase-1¹⁴. Of note, mitochondrial damage alone does not trigger NLRP3 signalling if priming is omitted⁹. It is not known, however, whether and how macrophage priming affects the mitochondrion and makes it more capable of producing ox-mtDNA to allow NLRP3 inflammasome activation¹³.

LPS induces macrophage mtDNA replication

The exposure of macrophages to NLRP3 activators results in mitochondrial damage, measurable by a drop in mitochondrial membrane potential and increased production of mitochondrial ROS (mtROS) (Extended Data Fig. 1a, b). Without previous priming, NLRP3 activators did not elicit extensive mtDNA oxidation or its cytoplasmic release (Extended Data Fig. 1c, d). Unlike NLRP3 activators, an AIM2 agonist, poly(dA-dT), did not induce mitochondrial damage, mtDNA oxidation or cytosolic mtDNA release (Extended Data Fig. 1a–d). These results suggest that macrophage priming affects the ability of mitochondria that were damaged after exposure to NLRP3 activators to produce ox-mtDNA and release mtDNA fragments to the cytosol. We and others have shown that mtDNA depletion by chronic treatment with low-dose ethidium bromide prevented NLRP3 inflammasome activation^{10,11}. To establish genetically the role of mtDNA in NLRP3 inflammasome

¹Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, School of Medicine, University of California San Diego, La Jolla, CA, USA. ²Department of Pathology, School of Medicine, University of California San Diego, La Jolla, CA, USA. ³Department of Medicine, School of Medicine, University of California San Diego, La Jolla, CA, USA. ⁴Department of Surgery, School of Medicine, University of California San Diego, La Jolla, CA, USA. ⁵Biomedical Translational Research Institute and the First Affiliated Hospital, Jinan University, Guangzhou, China. ⁶Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. ⁷Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ⁸Va Palo Alto Health Care System, Palo Alto, CA, USA. ⁹Department of Medicine, Division of Gastroenterology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁰Department of Medicine, Division of Endocrinology, Diabetes and Hypertension, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA. ¹¹These authors contributed equally: Zhenyu Zhong, Shuang Liang. *e-mail: karinoffice@ucsd.edu

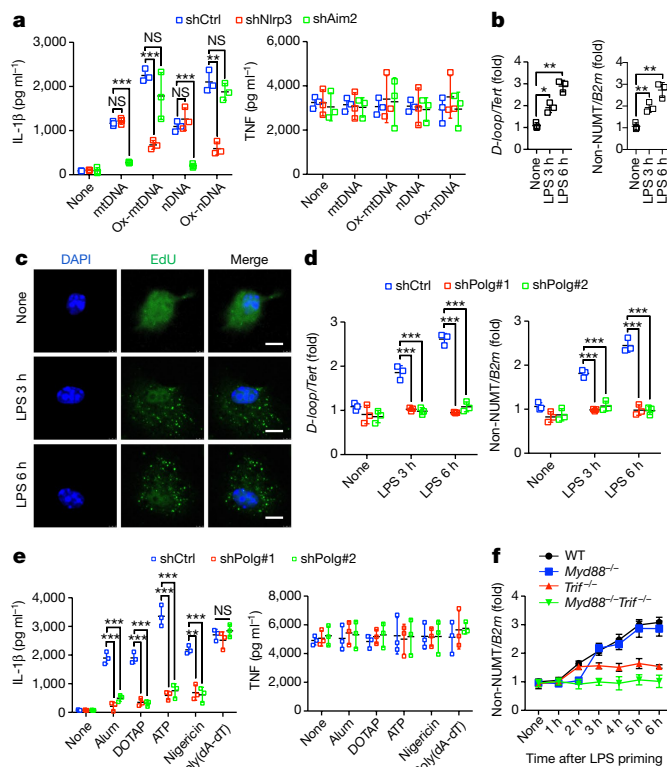


Fig. 1 | Newly synthesized mtDNA is required for NLRP3 inflammasome activation. **a**, *Tfam* Δ^{Mye} BMDMs transduced with shRNA against *Nlrp3*, *Aim2* (shNlrp3, shAim2) or control shRNA (shCtrl) and primed with LPS were incubated with synthetic mtDNA, ox-mtDNA, nuclear DNA (nDNA) or oxidized nuclear DNA (ox-nDNA), and the release of IL-1 β (left) and TNF (right) was measured 4 h later. **b**, Relative total mtDNA amounts were quantified by quantitative PCR (qPCR) with primers specific for the mitochondrial D-loop region or a region of mtDNA that is not inserted into nuclear DNA (non-NUMT) and primers specific for nDNA (*Tert*, *B2m*) in wild-type BMDMs before and after LPS priming. **c**, Fluorescent microscopy of EdU-labelled newly synthesized mtDNA in wild-type BMDMs before and after LPS priming. Scale bars, 5 μ m. Images are representative of three independent experiments. **d**, Relative total mtDNA amounts in wild-type BMDMs transduced with *Polg* shRNA (shPolg#1 and shPolg#2) or control shRNA, before and after LPS priming. **e**, IL-1 β (left) and TNF (right) release by shCtrl- or shPolg-transduced LPS-primed BMDMs treated with different inflammasome activators. **f**, Relative total mtDNA amounts in wild-type (WT), *Myd88*^{-/-} and *Trif*^{-/-} (*Trif* is also known as *Ticam1*) BMDMs after LPS priming. Data in **a**, **b**, **d**–**f** are mean \pm s.d. ($n = 3$ biological replicates). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired t -test. NS, not significant.

activation we crossed *LysM-cre* and *Tfam*^{fl/fl} mice¹⁵ to generate *Tfam* Δ^{Mye} mice that specifically lack TFAM (transcription factor A, mitochondrial) in myeloid cells (Extended Data Fig. 1e). TFAM binds mtDNA to promote its compaction and stabilization as well as replication and transcription¹⁶. *Tfam* ablation markedly reduced mtDNA content in mouse bone marrow-derived macrophages (BMDMs) (Extended Data Fig. 1f). *Tfam* Δ^{Mye} BMDMs did not produce mtROS and ox-mtDNA in response to NLRP3 activators and displayed defective caspase-1 activation and IL-1 β processing, while retaining normal pro-IL-1 β and NLRP3 induction, expression of general inflammasome components, normal AIM2 inflammasome activation and unaltered TNF expression (Extended Data Fig. 1g–l). To determine whether mtDNA oxidation is required for NLRP3 inflammasome activation, we incubated *Tfam* Δ^{Mye} BMDMs with hydrogen peroxide, the predominant ROS in activated macrophages, and assessed whether this rescues defective NLRP3 inflammasome activation. Notably, hydrogen peroxide, which induced IL-1 β release in *Tfam*^{fl/fl} BMDMs, failed to restore NLRP3

inflammasome activation in TFAM-deficient cells even in combination with nigericin (Extended Data Fig. 1m), suggesting that mtROS trigger NLRP3 inflammasome activation by promoting the production of ox-mtDNA. To further rule out the possibility that TFAM itself rather than mtDNA is required for NLRP3 inflammasome activation, we synthesized a 90-base-pair (bp) DNA fragment encompassing the D-loop (origin of mtDNA replication; also known as *mt-Dcr*) region of mouse mtDNA in the presence of the oxidized nucleotide 8-OH-dGTP. Transfection of synthetic D-loop ox-mtDNA into lipopolysaccharide (LPS)-primed *Tfam* Δ^{Mye} BMDMs induced IL-1 β production without any known NLRP3 activator (Fig. 1a). The effect was dependent on NLRP3, but not on AIM2 (Fig. 1a and Extended Data Fig. 2a), demonstrating that TFAM promotes NLRP3 inflammasome activation by facilitating ox-mtDNA formation or release. Transfection of oxidized nuclear DNA of the same length (90 bp) also resulted in IL-1 β production (Fig. 1a), indicating that NLRP3 detects specific nucleotide alteration(s) rather than DNA sequence or its cellular source. Indeed, replacement of 8-OH-dGTP with dGTP during in vitro DNA synthesis led to AIM2, but not NLRP3, inflammasome activation (Fig. 1a and Extended Data Fig. 2a). Neither form of DNA affected TNF synthesis (Fig. 1a). Cytosine methylation had no effect on the ability of DNA to activate inflammasome (Extended Data Fig. 2b). Finally, we confirmed that endogenous ox-mtDNA co-localizes with ASC-containing inflammasome complexes after NLRP3 activator treatment (Extended Data Fig. 2c).

Because TFAM is required for mtDNA replication and maintenance¹⁶, and priming is needed for ox-mtDNA production, we looked for a link between priming and mtDNA metabolism. Notably, macrophage stimulation with LPS resulted in a rapid and robust increase in mtDNA copy number, peaking at around 6 h and remaining increased for at least 24 h after stimulation (Fig. 1b and Extended Data Fig. 3a). Increased mtDNA abundance correlated with enhanced incorporation of 5-ethynyl-2'-deoxyuridine (EdU) into mitochondria-like cytoplasmic organelles (Fig. 1c). Because ablation of DNA polymerase- γ (*Polg*), the enzyme responsible for mtDNA replication¹⁷, prevented the LPS-induced increase in mtDNA abundance (Fig. 1d and Extended Data Figs. 2a, 3b, c), we reasoned that the increased mtDNA copy number in LPS-primed cells is due to new mtDNA synthesis and that the EdU-labelled organelles are indeed mitochondria, as their labelling was dependent on *Polg*. More importantly, blockade of mtDNA synthesis inhibited NLRP3 activator-induced IL-1 β production but had no effect on TNF synthesis (Fig. 1e). Although we cannot rule out a role for *Polg*-mediated DNA repair, we reason that repair of patches of damaged mtDNA alone is unlikely to cause a two- to threefold increase in mtDNA copy number. Although mtDNA replication is often accompanied by increased mitochondrial mass, mitochondrial residential proteins were not increased (Extended Data Fig. 3d). Nonetheless, LPS treatment may have stimulated mitochondrial fission (Extended Data Fig. 3e). These results suggest that LPS-induced mtDNA replication is an important signalling event in activated macrophages rather than serving a general homeostatic function.

As LPS binds TLR4, which signals via MyD88 and TRIF¹⁸, we examined the involvement of these adaptors in LPS-induced mtDNA replication. Although MyD88 was responsible for the initial increase in synthesis of mtDNA, TRIF took over at later time points (Fig. 1f). Both MyD88 (early) and TRIF (late) contributed to NLRP3 activator-induced mtDNA oxidation and the formation of ASC specks with which new mtDNA was co-localized (Extended Data Fig. 4a–c).

IRF1 controls mtDNA replication and NLRP3 activation

We searched for downstream targets of MyD88 and TRIF that could be involved in TLR-mediated stimulation of mtDNA synthesis and found that induction of interferon regulatory factor 1 (IRF1) followed similar kinetics to those of mtDNA synthesis and was similarly dependent on MyD88 at early time points and on TRIF at later stages (Extended Data Fig. 4d–f). Importantly, IRF1 ablation blocked the induction of

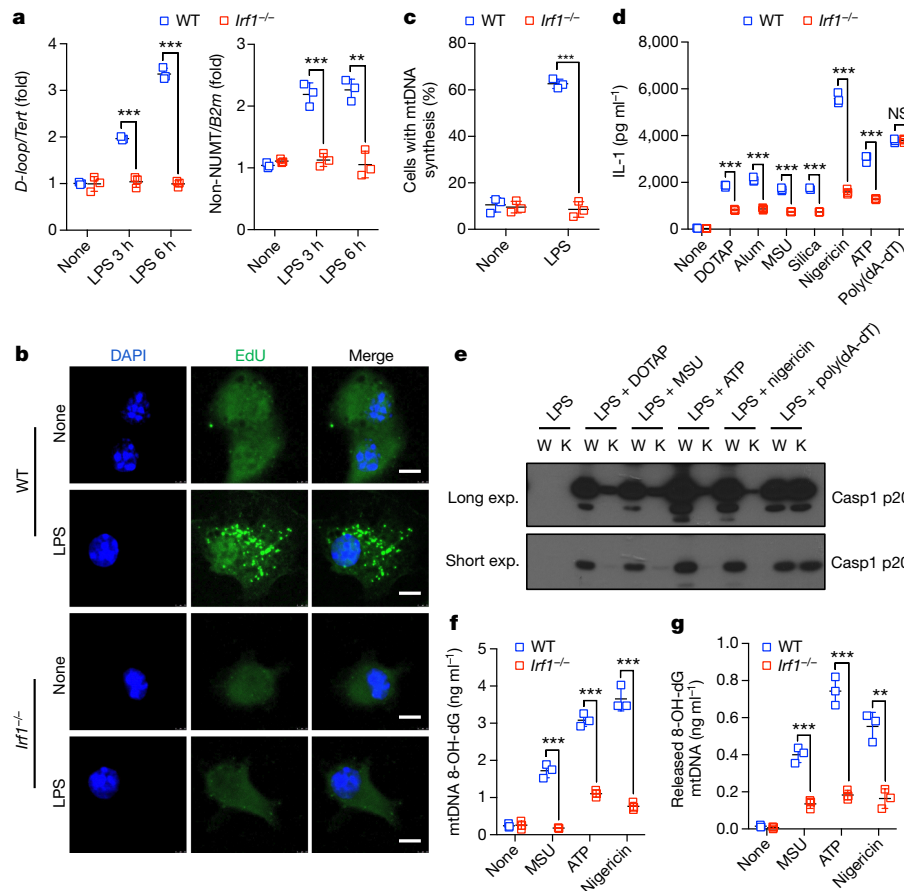


Fig. 2 | IRF1-dependent mtDNA synthesis, ox-mtDNA generation and NLRP3 inflammasome activation. **a**, Relative total mtDNA amounts in wild-type and *Irf1*^{-/-} BMDMs that were stimulated with LPS. **b**, Representative images showing EdU incorporation into mtDNA in wild-type and *Irf1*^{-/-} BMDMs incubated without or with LPS for 6 h. Scale bars, 5 μm. **c**, Percentages of cells undergoing mtDNA synthesis, as determined in **b** (*n* = 3 different microscopic fields per group; original magnification, ×40). **d**, IL-1β release by LPS-primed wild-type and *Irf1*^{-/-} BMDMs that were treated with different inflammasome activators.

mtDNA synthesis (Fig. 2a–c) and *Irf1*^{-/-} BMDMs showed a substantial reduction in NLRP3 activator-induced caspase-1 activation and IL-1β release, while retaining normal AIM2 inflammasome activation (Fig. 2d, e). The expression of pro-IL-1β and NLRP3 inflammasome components (including NEK7^{19–21}) and TNF secretion were also unaffected by IRF1 ablation, which also had no effect on NLRP3 activator-induced mitochondrial damage or mtROS production (Extended Data Fig. 5a–d). The IRF1 deficiency also did not affect caspase-11 mediated non-canonical inflammasome activation (Extended Data Fig. 5e, f). Owing to defective LPS-induced mtDNA replication, less ox-mtDNA was found in *Irf1*^{-/-} BMDMs (both in mitochondrial and cytosolic fractions) after NLRP3 activator challenge (Fig. 2f, g). IRF1 probably controls ox-mtDNA production and NLRP3 inflammasome activation through its effect on mtDNA replication.

CMPK2 controls mitochondrial DNA synthesis

Because IRF1 is a transcription factor, we searched for an IRF1 target gene that may be involved in mtDNA replication, and found that the IRF1 transcriptome²² included a gene coding for the mitochondrial deoxyribonucleotide kinase UMP-CMPK2 (hereafter referred to as CMPK2)²³. Of note, LPS priming resulted in strong induction of *Cmpk2* mRNA and protein that was IRF1 dependent and showed similar dependence on MyD88 and TRIF as IRF1 did (Fig. 3a and Extended Data Figs. 4d–f and 6a). As expected, newly synthesized CMPK2 entered mitochondria (Extended Data Fig. 6b). The 5' control

e, Immunoblot analysis of cleaved caspase-1 (Casp1 p20) in culture supernatants of wild-type and *Irf1*^{-/-} BMDMs that were treated with LPS plus different inflammasome activators. Data are representative of three independent experiments. exp., exposure; K, knockout (*Irf1*^{-/-}); W, wild type. **f**, **g**, Amounts of 8-OH-dG in mtDNA within mitochondria (**f**) or in the cytosol (**g**) of LPS-primed wild-type and *Irf1*^{-/-} BMDMs treated with NLRP3 activators. Data in **a**, **d**, **f** and **g** are mean ± s.d. (*n* = 3 biological replicates). ***P* < 0.01; ****P* < 0.001; two-sided unpaired *t*-test.

region of *Cmpk2* contains three IRF1-binding sites, the functionality of which was confirmed by chromatin immunoprecipitation (Extended Data Fig. 6c). CMPK2 is a mitochondrial nucleotide monophosphate kinase needed for salvage dNTP synthesis²³. Notably, other nucleoside/nucleotide kinases in this pathway and Polγ were not LPS-inducible (Extended Data Fig. 6d, e), suggesting that CMPK2 is the rate-limiting enzyme that controls the supply of dNTP precursors for LPS-induced mtDNA synthesis. CMPK2 phosphorylates dCMP to dCDP, which is further converted into dCTP by the constitutive deoxyribonucleotide diphosphate kinase NME4^{23,24}.

To validate the role of CMPK2 in LPS-induced mtDNA replication and NLRP3 inflammasome activation, we knocked down *Cmpk2* in wild-type BMDMs using short hairpin RNA (shRNA) (Extended Data Fig. 6f). CMPK2-deficient macrophages exhibited minimal NLRP3-dependent caspase-1 activation and IL-1β maturation relative to control CMPK2-sufficient cells, while retaining normal AIM2 responsiveness, TNF secretion and expression of NLRP3 inflammasome components as well as pro-IL-1β (Fig. 3b–d and Extended Data Fig. 7a). Although CMPK2 ablation did not affect mitochondrial damage or mtROS production after NLRP3 activator exposure (Extended Data Fig. 7b, c), CMPK2-deficient BMDMs did not upregulate mtDNA synthesis after LPS stimulation and barely produced ox-mtDNA (Fig. 3e–g and Extended Data Fig. 7d, e). Two other TLR ligands, the TLR2 agonist Pam3CSK and the TLR3 agonist polyriboinosinic:polyribocytidylic acid (poly(I:C)), which can prime inflammasome activation,

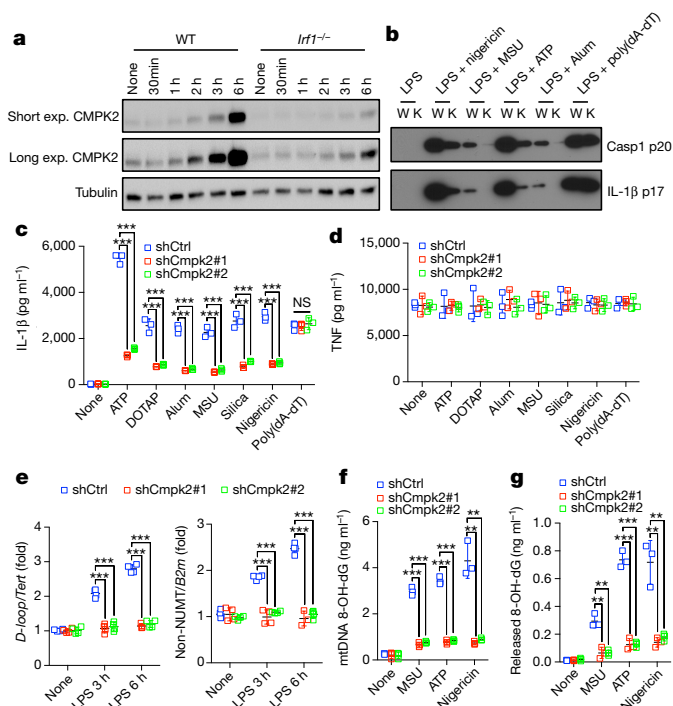


Fig. 3 | CMPK2 controls mtDNA synthesis, ox-mtDNA generation and NLRP3 inflammasome activation. **a**, Time-course analysis of CMPK2 accumulation in wild-type and *Irfl*^{-/-} BMDMs before and after LPS priming. **b**, Immunoblot analysis of cleaved caspase-1 (Casp1 p20) and mature IL-1β (p17) in the supernatants of shCtrl (wild-type, W) and shCmpk2 (knockdown, K) BMDMs that were stimulated with LPS plus the indicated inflammasome activators. Data in **a** and **b** represent three independent experiments. **c**, **d**, IL-1β (**c**) and TNF (**d**) secretion by LPS-primed BMDMs transduced with control shRNA or *Cmpk2* shRNA (shCmpk2#1 and shCmpk2#2), and incubated with different inflammasome activators. **e**, Relative total mtDNA amounts in control and *Cmpk2* shRNA BMDMs before and after LPS stimulation. **f**, **g**, Amounts of 8-OH-dG in mtDNA within mitochondria (**f**) or in the cytosol (**g**) of LPS-primed BMDMs transduced with control shRNA or *Cmpk2* shRNA that were treated with different NLRP3 activators. Data in **c–g** are mean ± s.d. ($n = 3$ biological replicates, except for $n = 4$ in **e**). ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired *t*-test.

also induced mtDNA replication via the IRF1 and CMPK2 pathway (Extended Data Fig. 7f, g), indicating that this pathway represents a common mechanism by which macrophages upregulate mtDNA abundance in response to TLR stimulation. NME4 ablation (Extended Data Fig. 6f) also blocked LPS-induced mtDNA synthesis, and reduced ox-mtDNA production as well as IL-1β secretion after NLRP3 inflammasome activation, without affecting AIM2 inflammasome activation or TNF production (Extended Data Fig. 8a–e). By contrast, increasing the cellular dNTP pool by the ablation of *Samhd1*, the dominant nucleotide triphosphate hydrolase²⁵, the expression of which is induced by LPS in a TRIF-dependent manner (Extended Data Fig. 8f), enhanced new mtDNA synthesis and augmented IL-1β secretion after NLRP3 activator stimulation (Extended Data Fig. 8g–j). Enhanced IL-1β production in *Samhd1*^{-/-} BMDMs remained dependent on mtDNA synthesis.

NLRP3 activation depends on CMPK2 catalytic activity

To confirm that CMPK2 promotes NLRP3 inflammasome activation by providing dCTP for mtDNA synthesis, we generated a catalytically inactive CMPK2 variant, CMPK2(D330A), by replacing the highly conserved aspartate (D) residue in its catalytic pocket^{23,26} with alanine (A). Expression of wild-type CMPK2 in *Irfl*^{-/-} BMDMs restored LPS-stimulated mtDNA replication but did not enhance it in wild-type BMDMs (Fig. 4a). Although CMPK2 reconstitution did not

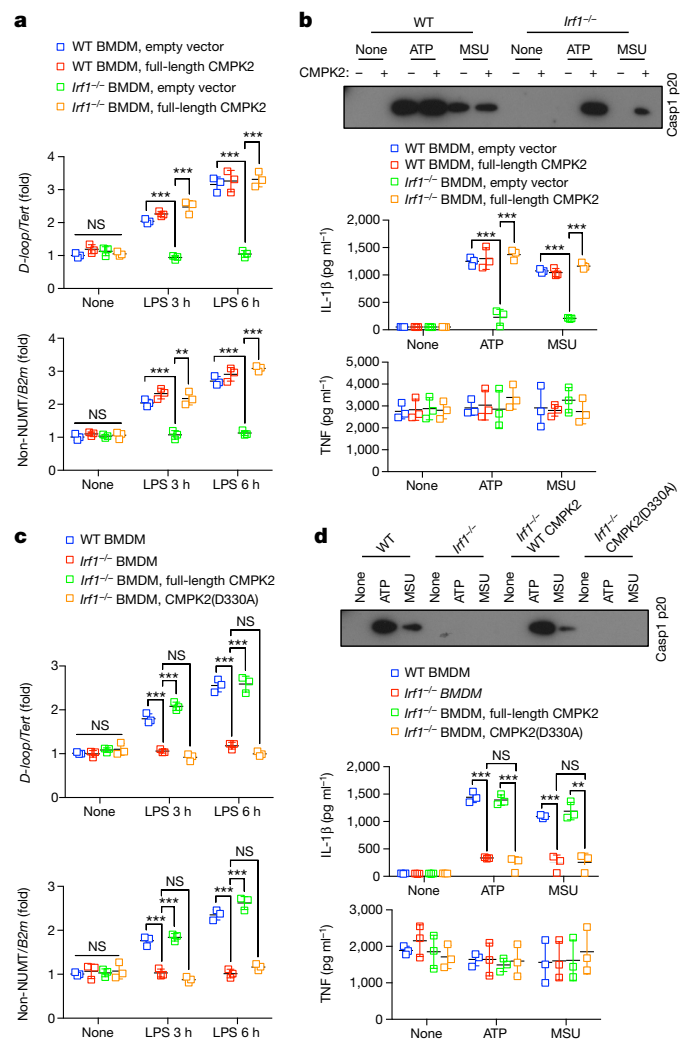


Fig. 4 | CMPK2 catalytic activity is required for NLRP3 inflammasome activation. **a**, Relative total mtDNA amounts in LPS-treated wild-type and *Irfl*^{-/-} BMDMs transduced with CMPK2-encoding or empty lentiviruses. **b**, Top, immunoblot analysis of Casp1 p20 in the supernatants of CMPK2 lentivirus-transduced LPS-primed wild-type and *Irfl*^{-/-} BMDMs that were stimulated with the indicated NLRP3 activators. Data are representative of three independent experiments. Bottom, IL-1β and TNF secretion by the above cells. **c**, Relative total mtDNA amounts in LPS-treated wild-type and *Irfl*^{-/-} BMDMs transduced with either wild-type CMPK2- or CMPK2(D330A)-encoding lentiviruses. **d**, Top, immunoblot analysis of Casp1 p20 in supernatants of wild-type CMPK2- or CMPK2(D330A)-encoding lentiviruses transduced LPS-primed wild-type and *Irfl*^{-/-} BMDMs that were stimulated with NLRP3 activators. Data are representative of three independent experiments. Bottom, IL-1β and TNF secretion by the above cells. Data in **a**, **b** (bottom), **c** and **d** (bottom) are mean ± s.d. ($n = 3$ biological replicates). ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired *t*-test.

alter the expression of pro-IL-1β, NLRP3, ASC and pro-caspase-1, and had no effect on mitochondrial damage or ROS induction by NLRP3 activators (Extended Data Fig. 9a–c), it restored NLRP3 inflammasome activation (Fig. 4b). By contrast, re-expression of CMPK2(D330A) did not restore LPS-induced mtDNA synthesis or NLRP3 inflammasome activation (Fig. 4c, d and Extended Data Fig. 9d). These results strongly support the notion that the induction of new mtDNA replication, which depends on CMPK2 catalytic activity, is required for the production of ox-mtDNA by mitochondria that have been damaged by exposure to NLRP3 activators, with ox-mtDNA being responsible for subsequent NLRP3 inflammasome activation.

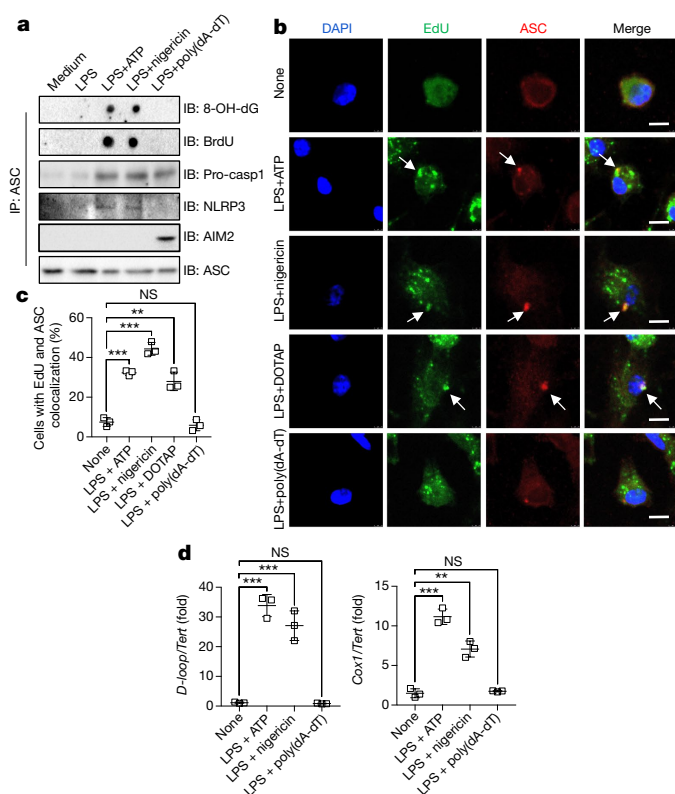


Fig. 5 | Newly synthesized mtDNA associates with the NLRP3 inflammasome complex. **a**, Inflammasomes from BMDMs after indicated treatments were immunoprecipitated with ASC antibodies. The immunocomplexes were spotted on a nitrocellulose membrane, UV-crosslinked and probed with antibodies to 8-OH-dG and BrdU, or separated by SDS-PAGE and immunoblotted with antibodies to pro-Casp1, NLRP3, AIM2 and ASC. Data are representative of three independent experiments. **b**, Representative fluorescence microscopy images of EdU-labelled BMDMs that were co-stained for ASC, EdU and DAPI before and after stimulation with LPS plus different inflammasome activators. Arrows indicate co-localization of EdU and ASC signals. Scale bars, 5 μ m. **c**, Percentages of cells with ASC and EdU co-localization as determined in **b**. Data are mean \pm s.d. (n = 3 different microscopic fields per group; original magnification, $\times 40$). **d**, Relative D-loop (left) and Cox1 (right) mtDNA amounts in inflammasome complexes isolated as in **a**. Data are mean \pm s.d. (n = 3 biological replicates). **P < 0.01; ***P < 0.001; two-sided unpaired t-test.

New mtDNA binds NLRP3 after mitochondrial damage

To verify the role of new mtDNA synthesis in NLRP3 signalling, we incubated BMDMs with bromodeoxyuridine (BrdU) to label newly synthesized mtDNA, stimulated these cells with LPS and ATP or LPS and nigericin, and immunoprecipitated inflammasome complexes with ASC antibodies. The resulting immunocomplexes contained NLRP3, BrdU-labelled DNA, and 8-OH-dG (Fig. 5a). However, ASC immunocomplexes isolated from BMDMs that were stimulated with an AIM2 agonist did not contain NLRP3, BrdU-labelled DNA or 8-OH-dG (Fig. 5a), confirming specific interaction between newly synthesized and oxidized DNA and the NLRP3 inflammasome complex. To visualize this interaction, we examined the co-localization of ASC-containing inflammasome aggregates and EdU-labelled mtDNA before and after NLRP3 activator treatment. Remarkably, ATP, nigericin or 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) together with LPS induced co-localization of newly synthesized mtDNA with ASC specks, whereas the AIM2 agonist failed to do so, even in combination with LPS, which still induced EdU incorporation into cytoplasmic organelles (Fig. 5b, c). To confirm that NLRP3 inflammasome-associated DNA was of mitochondrial origin, we extracted DNA from ASC immunocomplexes and subjected it to PCR amplification. This resulted in detection of both D-loop and cytochrome

c oxidase subunit 1 (COX1) mitochondrial sequences in NLRP3-, but not AIM2-, inflammasome complexes (Fig. 5d). Notably, the fold-enrichment of D-loop mtDNA was higher than that of Cox1 (also known as mt-Co1) mtDNA (Fig. 5d), suggesting that mtDNA synthesis, which originates at the D-loop, is responsible for the generation of ox-mtDNA that binds NLRP3.

IRF1 controls in vivo NLRP3 activation

Lastly, we established a requirement for IRF1 in NLRP3 inflammasome activation in vivo. Intraperitoneal LPS injection is sufficient for the induction of an IL-1 β - and NLRP3-dependent acute systemic inflammation that eventually leads to death^{5,27}. Relative to wild-type mice, *Irf1*^{-/-} mice exhibited markedly reduced IL-1 β secretion but little change in TNF production and were largely resistant to LPS-induced death (Extended Data Fig. 10a–c). Importantly, *Irf1*^{-/-} peritoneal macrophages isolated 3 h after LPS injection exhibited a lower mtDNA copy number than wild-type macrophages (Extended Data Fig. 10d). *Irf1*^{-/-} mice were also refractory to alum-induced NLRP3 inflammasome-dependent IL-1 β production, and exhibited reduced neutrophil and monocyte infiltration relative to wild-type counterparts (Extended Data Fig. 10e–g).

Discussion

The NLRP3 inflammasome has a central role in numerous acute and chronic inflammatory and degenerative diseases¹, but the mechanism that controls its activation is poorly understood. The difficulties stem from the fact that the NLRP3 inflammasome is activated by structurally diverse and chemically unrelated entities, none of which binds NLRP3 itself²⁸. NLRP3 inflammasome signalling depends on priming and activation, but whether and how priming affects inflammasome assembly and subsequent activation has remained elusive. Although mitochondrial damage and mtROS production were shown to be essential for NLRP3 inflammasome activation^{9,12}, they are induced by NLRP3 activators even in non-primed macrophages, where they do not result in inflammasome activation. We now show (Extended Data Fig. 10h) that priming and activation are coupled through the induction of new mtDNA synthesis, a hitherto unrecognized component of NLRP3 signalling. TLR4 engagement triggers MyD88/TRIF-dependent signalling that activates IRF1 to ultimately induce the expression of CMPK2, a mitochondrial nucleotide kinase with activity that is rate-limiting for de novo mtDNA synthesis. As a highly conserved enzyme, CMPK2 catalyses the synthesis of dCDP, which is further converted to dCTP by constitutively expressed NME4, thereby supplying an essential dNTP for mtDNA synthesis. Curiously, another TRIF-induced gene, *Samhd1*, encodes a dNTP hydrolase that curtails mtDNA synthesis and IL-1 β production. Although new mtDNA synthesis is dispensable for mitochondrial damage and mtROS production, it is needed for the generation of what may be the ultimate NLRP3 ligand: ox-mtDNA. We demonstrate that newly replicated mtDNA co-precipitates and co-localizes with NLRP3 inflammasome complexes in macrophages incubated with LPS and NLRP3 activators. We postulate that newly synthesized mtDNA, which is yet to be packaged into a highly condensed nucleoid structure by TFAM¹⁶, is highly susceptible to oxidation and nuclease action, resulting in the production of ox-mtDNA fragments that are released via membrane pores that open after exposure to NLRP3 activators.

Mitochondrial involvement in NLRP3 inflammasome activation was first proposed by Tschopp and colleagues, but remained debatable¹². Subsequently, Arditi and co-workers demonstrated that ox-mtDNA generated during apoptosis can bind NLRP3¹³. We previously showed that autophagic elimination of mitochondria that were damaged after macrophage exposure to NLRP3 activators attenuates NLRP3 inflammasome activation¹¹. Our current results show that long-term inhibition of mtDNA synthesis via TFAM ablation, or specific interference with TLR-induced mtDNA synthesis through ablation or inactivation of IRF1, CMPK2, NME4 or Pol γ , prevents NLRP3 inflammasome activation. The latter, however, can be fully restored by the re-introduction

of oxidized DNA into macrophages in the absence of any NLRP3 activators or apoptotic stimuli. Because any oxidized DNA regardless of its cellular source or sequence can activate the NLRP3 inflammasome, we propose that NLRP3 is likely to recognize the presence of 8-OH-dG on the oxidized DNA molecule.

Lastly, we can only speculate why TLR engagement, which leads to MyD88/TRIF-dependent and IRF1-mediated CMPK2 induction, stimulates mtDNA synthesis. Perhaps mtDNA replication is needed for the maintenance of proper mitochondrial function to meet the energy and/or signalling demands of activated macrophages that are actively engaged in phagocytosis and other immune functions. Alternatively, mtROS production may result in mitochondrial damage and the elimination of damaged mitochondria via mitophagy. The induction of new mtDNA synthesis would allow macrophages to cope with TLR-stimulated ROS production²⁹ that may otherwise result in mitochondrial depletion. A more intriguing possibility raised by the evolutionary relationships between mitochondria and intracellular parasitic bacteria is that CMPK2 induction may be an evolutionary relic that once helped such bacteria to survive and replicate within activated phagocytes²³. Because CMPK2 catalytic activity is essential for NLRP3 inflammasome activation, our results outline an entirely new approach for inhibiting NLRP3-dependent immunopathologies.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0372-z>.

Received: 4 August 2017; Accepted: 13 June 2018;

Published online 25 July 2018.

- Gross, O., Thomas, C. J., Guarda, G. & Tschopp, J. The inflammasome: an integrated view. *Immunol. Rev.* **243**, 136–151 (2011).
- Kotas, M. E. & Medzhitov, R. Homeostasis, inflammation, and disease susceptibility. *Cell* **160**, 816–827 (2015).
- Karin, M. & Clevers, H. Replicative inflammation takes charge of tissue regeneration. *Nature* **529**, 307–315 (2016).
- Zhong, Z., Sanchez-Lopez, E. & Karin, M. Autophagy, inflammation, and immunity: a troika governing cancer and its treatment. *Cell* **166**, 288–298 (2016).
- Lu, A. et al. Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **156**, 1193–1206 (2014).
- Heneka, M. T., Kummer, M. P. & Latz, E. Innate immune activation in neurodegenerative disease. *Nat. Rev. Immunol.* **14**, 463–477 (2014).
- Lamkanfi, M. & Dixit, V. M. Inflammasomes and their roles in health and disease. *Annu. Rev. Cell Dev. Biol.* **28**, 137–161 (2012).
- Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
- Latz, E., Xiao, T. S. & Stutz, A. Activation and regulation of the inflammasomes. *Nat. Rev. Immunol.* **13**, 397–411 (2013).
- Nakahira, K. et al. Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nat. Immunol.* **12**, 222–230 (2011).
- Zhong, Z. et al. NF- κ B restricts inflammasome activation via elimination of damaged mitochondria. *Cell* **164**, 896–910 (2016).
- Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221–225 (2011).
- Shimada, K. et al. Oxidized mitochondrial DNA activates the NLRP3 inflammasome during apoptosis. *Immunity* **36**, 401–414 (2012).
- Jiang, X. & Wang, X. Cytochrome c-mediated apoptosis. *Annu. Rev. Biochem.* **73**, 87–106 (2004).
- Hamanaka, R. B. et al. Mitochondrial reactive oxygen species promote epidermal differentiation and hair follicle development. *Sci. Signal.* **6**, ra8 (2013).
- Kang, D., Kim, S. H. & Hamasaki, N. Mitochondrial transcription factor A (TFAM): roles in maintenance of mtDNA and cellular functions. *Mitochondrion* **7**, 39–44 (2007).
- He, Y., Zeng, M. Y., Yang, D., Motro, B. & Núñez, G. NEK7 is an essential mediator of NLRP3 activation downstream of potassium efflux. *Nature* **530**, 354–357 (2016).
- Shi, H. et al. NLRP3 activation and mitosis are mutually exclusive events coordinated by NEK7, a new inflammasome component. *Nat. Immunol.* **17**, 250–258 (2016).
- Schmid-Burgk, J. L. et al. A genome-wide CRISPR (clustered regularly interspaced short palindromic repeats) screen identifies NEK7 as an essential component of NLRP3 inflammasome activation. *J. Biol. Chem.* **291**, 103–109 (2016).
- Jehl, S. P., Nogueira, C. V., Zhang, X. & Starnbach, M. N. IFN γ inhibits the cytosolic replication of *Shigella flexneri* via the cytoplasmic RNA sensor RIG-I. *PLoS Pathog.* **8**, e1002809 (2012).
- Xu, Y., Johansson, M. & Karlsson, A. Human UMP-CMP kinase 2, a novel nucleoside monophosphate kinase localized in mitochondria. *J. Biol. Chem.* **283**, 1563–1571 (2008).
- Milon, L. et al. The human nm23-H4 gene product is a mitochondrial nucleoside diphosphate kinase. *J. Biol. Chem.* **275**, 14264–14272 (2000).
- Chen, Y. L., Lin, D. W. & Chang, Z. F. Identification of a putative human mitochondrial thymidine monophosphate kinase associated with monocytic/macrophage terminal differentiation. *Genes Cells* **13**, 679–689 (2008).
- Martinson, F., Mayor, A. & Tschopp, J. The inflammasomes: guardians of the body. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
- Elliott, E. I. & Sutterwala, F. S. Initiation and perpetuation of NLRP3 inflammasome activation and assembly. *Immunol. Rev.* **265**, 35–52 (2015).
- West, A. P. et al. TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* **472**, 476–480 (2011).

Acknowledgements We thank eBioscience, Cell Signaling Technologies, Santa Cruz Technologies, and Thermo Fisher for gifts of reagents, and N. Yan and J. Rehwinkel for SAMHD1-deficient murine bone marrow. Z.Z. was supported by Cancer Research Institute (CRI) Irvington Fellowship, Prevent Cancer Foundation Board of Directors Research Fund, and American Association for the Study of Liver Diseases Pinnacle Research Award; F.H. was supported by Eli Lilly LIFA program; S.S. was supported by fellowships from CRI-Irvington and Prostate Cancer Foundation; J.W. was supported by a Canadian Institutes of Health Research fellowship (MFE-135425). Research was supported by NIH grants AI043477 and CA163798 to M.K., AA020172 and DK085252 to E.S., ES010337 to M.K. and E.S., DK109724 and P30DK063491 to A.L.H., DK099205, DK101737 and DK111866 to T.K., AA022614 to T.K. and M.K., Leukemia and Lymphoma Society SCOR grant 20132569 to T. Kipps and M.K., and the Alliance for Lupus Research grant 257214 and CART Foundation to M.K., who is an American Cancer Research Society Professor and holds the Ben and Wanda Hildyard Chair for Mitochondrial and Metabolic Diseases.

Reviewer information Nature thanks M. Ardit, M. Murphy and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Z.Z., S.L. and M.K. conceived the project. Z.Z. and S.L. designed and performed most of the experiments. E.S.-L., F.H., S.S., X.-J.L., J.W. and S.D. provided technical assistance. E.S., B.S., A.L.H., T.K. and H.G. provided reagents and research tools. Z.Z., S.L. and M.K. wrote the manuscript with input from all authors.

Competing interests The University of California San Diego is in the process of applying for a patent application (US Provisional Application Serial no. 62/690,175) covering the use of IRF1 and/or CMPK2 genetic/chemical inhibitors to treat NLRP3 inflammasome-associated diseases that lists Z.Z. and M.K. as inventors.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0372-z>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0372-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to M.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized except for the in vivo studies in which the age- and gender-matched mice were randomly allocated to different experimental groups based on their genotypes. Investigators were not blinded to allocation during experiments and outcome assessment except for microscopic analysis of immunofluorescent staining results.

Mice. C57BL/6, *Irf1*^{-/-}, *LysM-cre* and *Tfam*^{fl/fl} mice in the C57BL/6 background were purchased from Jackson Laboratories. *LysM-cre* mice were crossed with *Tfam*^{fl/fl} mice to generate *Tfam*^{ΔMyc} mice. *Myd88*^{-/-}, *Trif*^{-/-} and *Myd88*^{-/-} *Trif*^{-/-} mice were previously described³⁰. All mice were bred and maintained at the University of California San Diego (UCSD) and were treated in accordance with guidelines of the Institutional Animal Care and Use Committee.

Reagents. Ultrapure LPS (*E. coli* O111:B4) was from Invivogen. Silica and ATP were from Sigma-Aldrich. Inject Alum and streptavidin-horseradish peroxidase (HRP) were from Pierce. MitoSOX, Click-IT EDU microplate kit and TSA kit (12) were from Life Technologies. Monosodium urate crystal (MSU) was from Enzo Life Science. TMRM was from AnaSpec Inc. (CA94555). DOTAP liposomes were made by Encapsula NanoSciences as described previously³¹. 90-bp fragments of nDNA and mtDNA with or without oxidation (8-OH-dGTP) or methylation (5-Me-dCTP) were from BioSynthesis. L929 cells (from ATCC) were authenticated before delivery to our laboratory, and were routinely tested negative for mycoplasma contamination. Antibodies used for immunoblot analysis were: anti-mouse IL-1β (12426S, Cell Signaling Technologies), anti-mouse NLRP3 (AG-20B-0014-C100, Adipogen), anti-mouse AIM2 (sc-137967, Santa Cruz Biotechnology), anti-mouse ASC (AG-25b-0006-C100, Adipogen), anti-mouse caspase-1 (AG-20B-0042-C100, Adipogen), anti-mouse deoxyguanosine kinase (ab38013, Abcam), anti-mouse thymidine kinase 2 (ab38302, Abcam), anti-mouse AK2 (ab166901, Abcam), anti-mouse NME4 (LS-C409886, LifeSpan BioSciences), anti-mouse TFAM (ab131607, Abcam), anti-mouse NEK7 (ab133514, Abcam), anti-mouse ATP5B (MAB3494, Millipore), rabbit anti-8-OH-dG (bs-1278R, Bioss), mouse anti-8-OH-dG (200-301-A99, Rockland), anti-BrdU (B8434, Sigma-Aldrich) and anti-tubulin (T5168, Sigma-Aldrich).

Macrophage culture and stimulation. Primary BMDMs were generated by culturing mouse bone marrow cells in the presence of 20% (v/v) L929 conditional medium for 7 days as described³². BMDMs were seeded in 6-, 24- or 48-well plates overnight in FBS-free DMEM medium. On day 2, after priming with ultrapure LPS (200 ng ml⁻¹) for 4 h, BMDMs (1 × 10⁶ cells ml⁻¹) were further stimulated with ATP (4 mM) or nigericin (10 μM) for 45 min unless otherwise indicated or DOTAP liposomes (100 μg ml⁻¹), alum (500 μg ml⁻¹), silica (600 μg ml⁻¹) and MSU (600 μg ml⁻¹) for 4 h. To activate the AIM2 inflammasome, macrophages were primed with LPS as above, followed by transfection of poly(dA-dT) (1 μg ml⁻¹) using Lipofectamine 2000 (Life Technologies) according to manufacturer's protocol. Similar approaches were used to transfect BMDMs with synthetic nuclear or mitochondrial DNA fragments. To activate caspase-11 non-canonical inflammasome, LPS was delivered into BMDMs using FuGENE HD transfection reagent (Promega) according to manufacturer's instruction. Culture supernatants were collected 4 h after infection and IL-1β release was measured by ELISA. Supernatants and cell lysates were collected for ELISA and immunoblot analyses. Knockdown of *Cmpk2*, *Nme4*, *Nlrp3* or *Aim2* was done by lentiviral transduction of primary BMDMs as described previously¹¹. Sequences of specific shRNAs (from Sigma shRNA Mission library) used in this study are as follows: shCmpk2#1 (5'-CCGGGTTTCGTCAGAAAGGTGGAAATCTCGA GATTTCCAACTTCTGACGAAACTTTT-3'); shCmpk2#2 (5'-CCG GTCTGCTTAACCTTCTGCGGTGTTCTCGAGAACCCGACAGATTAAGCAG ATTTT-3'); shNme4#1 (5'-CCGGCAGTGTTCACATCAGCAGAACTCGA GTTCTCTGCTGATGTGAACACTGTTT-3'); shNme4#2 (5'-CCGG CCTCTGTCAACAAGAAGTCAACTCGAGTTGACTTCTTGTGACAGAGG TTTT-3'); shPolg#1 (5'-CCGGCGGACCTTATAATGATGTGAACCTCGA GTTCAATCATATAAGTCCGTTTGTG-3'); shPolg#2 (CCGGCGATACAT ATGAGCATGCACATCTCGAGATGTGCATGCTCATAGTATCGTTT-3'); shNlrp3#1 (5'-CCGGCCATACCTTCAGTCTTGTCTCTCGAGAAGACA AGACTGAAGGTATGGTTT-3'); shNlrp3#2 (5'-CCGGCCGCGCTTA CTTCATCTGTTCTCGAGAACAGATGAAGTAAGGCGGTTT-3'); shAim2#1 (5'-CCGGGCTTGTCTTAAGGCTTGGATCTCGAG ATCCCAAGCCTTAGACAAAGCTTTT-3'); shAim2#2 (5'-CCGGGCCATGT GGAACAATTGTGAACCTCGAGTTCACAATTGTTCCATATGGCTTTT-3').

RNA isolation and qPCR. RNA was isolated from BMDMs and reverse transcribed, and qPCR was performed as previously described³¹. Primer sequences are as follows. *Irf1* F: 5'-AATTCCAACCAATCCAGG-3'; *Irf1* R: 5'-AGGCATCCTTGTGTATGTC-3'; *Cmpk2* F: 5'-GGCAATTATCTCGT GGCTTC-3'; *Cmpk2* R: 5'-GTAGCTATGGCGTAGGTGGC-3'; *dGK* F: 5'-TCTGCATTGAAGGCAACATC-3'; *dGK* R: 5'-CTGCCACGCTGCT ATAGGTT-3'; *Ak2* F: 5'-AGATTCCGAAGGGCATCC-3'; *Ak2* R: 5'-GGC

CAATGACAGACACAAA-3'; *Tk2* F: 5'-TCACCTGTACGGTTGATGGA-3'; *Tk2* R: 5'-GAATCGCGTAGTCAACCTCG-3'; *Nme4* F: 5'-GGACACACC GACTCAACAGA-3'; *Nme4* R: 5'-CACAGAATCGCTAGCATGGA-3'; *Polg* F: 5'-ACGTGGAGGTCTGCTTGG-3'; *Polg* R: 5'-AGTAACGCT CTTCCACACAGC-3'; *Hprt1* F: 5'-CTGGTGAAGGACCTCTCG-3'; *Hprt1* R: 5'-TGAAGTACTCATATAGTCAAGGGCA-3'.

ELISA. Paired (capture and detection) antibodies and standard recombinant mouse IL-1β (from R&D Systems) and TNF (from eBioscience) were used to determine cytokine concentrations in cell culture supernatants and mouse sera according to manufacturer's instructions.

Measurement of total mtDNA. Macrophages were primed with LPS (200 ng ml⁻¹) for the indicated time. Total DNA was isolated using Allprep DNA/RNA Mini Kit (catalogue 80204, Qiagen) according to manufacturer's instruction. mtDNA was quantified by qPCR using primers specific for the mitochondrial D-loop region or a specific region of mtDNA that is not inserted into nuclear DNA (non-NUMT)³³. Nuclear DNA encoding *Tert* and *B2m* was used for normalization. Primer sequences are as follows: *D-loop* F: 5'-AATCTACCATCCTCCGTGAACCC-3'; *D-loop* R: 5'-TCAGTTTGTAGTACCCCAAGTTTAA-3'; *Tert* F: 5'-CTAGCT CATGTGTCAAGACCCTCTT-3'; *Tert* R: 5'-GCCAGCAGCTTCTCTCTGTT-3'; *B2m* F: 5'-ATGGGAAGCCGAACATACTG-3'; *B2m* R: 5'-CAGTCTCAGTGGG GGTGAAT-3'; non-NUMT F: 5'-CTAGAAACCCGGAACCAAAA-3', and non-NUMT R: 5'-CCAGCTATCACCAAGCTCGT-3'.

Chromatin immunoprecipitation. Chromatin immunoprecipitation was performed using Pierce Agarose ChIP Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. In brief, wild-type and *Irf1*^{-/-} primary BMDMs were treated with or without LPS, and then crosslinked with formaldehyde to generate DNA-protein cross-links. Cell lysates were digested with micrococcal nuclease to generate chromatin fragments and then subjected to immunoprecipitation with IRF-1 antibody or IgG isotype control. The immunoprecipitated chromatinized DNA was recovered and purified, followed by qPCR amplification with primers flanking the IRF-1 binding sites of the *Cmpk2* promoter. In all experiments, the *Cxcl10* promoter region that is known to include IRF-1 binding sites was used as a positive control.

Inflammasome immunoprecipitation. Wild-type BMDMs were primed with LPS (200 ng ml⁻¹) for 6 h in the presence of BrdU (10 μM) followed by treatment with ATP or nigericin for 60 min. The cells were washed twice with PBS and immunoprecipitation was performed using the Pierce Classic Magnetic IP/Co-IP Kit (Thermo Fisher Scientific) per manufacturer's instructions. In brief, cells were collected as described above and lysed in lysis buffer supplemented with Protease Inhibitor Cocktail (Life Technologies), incubated with rabbit anti-ASC polyclonal antibody (AG-25b-0006-C100, Adipogen), rotated overnight at 4°C. Magnetic beads were added and incubated with lysates on a rotator for 1 h at room temperature. Beads were then washed, bound fractions were eluted with non-reducing sample buffer and supplemented with dithiothreitol (DTT). For detection of BrdU and 8-OH-dG in the ASC immunoprecipitation products, the eluted samples were dot-blotted and UV cross-linked to a nitrocellulose membrane that was immunoblotted with BrdU monoclonal antibody (BU33; Sigma) or 8OH-dG BrdU monoclonal antibody (15A3; Rockland Immunochemicals). For the detection of NLRP3, AIM2, ASC and pro-caspase-1 in the ASC immunoprecipitation products, the eluted samples were heated to 95°C for 5 min, the gel was separated by SDS-PAGE, transferred to a nitrocellulose membrane and immunoblotted with antibodies against ASC (2EI-7, Millipore), NLRP3 (AG-20B-0014-C100, Adipogen), pro-casp1 (AG-20B-0042-C100, Adipogen) and AIM2 (sc-137967, Santa Cruz Biotechnology). For the detection of mtDNA in the ASC immunoprecipitation products, DNA was extracted from eluted samples and qPCR was performed to amplify the *D-loop* region or the mitochondrial gene encoding cytochrome c oxidase 1 (*Cox1*). Nuclear DNA encoding *Tert* was used for normalization. Primer sequences are as follows: *D-loop* and *Tert* primer sequences are as described above; *Cox1* F: 5'-GCCCCAGATATAGCATTCCC-3', and *Cox1* R: 5'-GTTTCATCTGTTCTCTGCTCC-3'.

Immunoblot analysis. Cells were lysed in RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) containing a protease inhibitor cocktail (Roche, 11836153001) and a phosphatase inhibitor cocktail (Sigma-Aldrich, P5726). Protein concentrations were quantified using BCA Protein Assay Kit (Pierce, 23225). Equal amounts of protein were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were then incubated with antibodies against IRF1, CMPK2, TFAM, NEK7, AK2, dGK, NME4, TK2, Polγ, NLRP3, ASC, α-tubulin, caspase-1 or IL-1β (as described above), followed by incubation with the appropriate secondary HRP-conjugated antibodies, and development with ECL.

Measurement of mitochondrial membrane potential and mtROS. Mitochondrial membrane potential (Ψ_m) and mtROS were measured as previously described¹¹ using TMRM and MitoSOX, respectively. In brief, BMDMs were put onto 6-well plates and primed with LPS (200 ng ml⁻¹) followed by treatment with ATP or

nigericin for 30 min or alum, silica, MSU or DOTAP liposomes for 3 h, after which the cells were washed twice with PBS. Then cells were detached and transferred into sterile 1.7-ml tubes for NLRP3 activator treatment. All the cell-staining procedures after NLRP3 activator stimulation were then performed in the tube. For Ψ_m measurement, BMDMs were loaded with 200 nM TMRM for 30 min and washed twice with 50 nM TMRM. For mtROS measurement, BMDM were loaded with 4 μ M MitoSOX for 20 min and washed twice with PBS. After staining and washing, cells were resuspended in PBS and counted. Equal numbers of cells from different treatment groups were then plated onto 96-well plates for fluorescence reading to minimize the variation due to unequal cell numbers or differences in the cell attachment among treatment groups. Fluorescence intensity was determined using a FilterMax F5 multimode plate reader (Molecular Devices), and the data were normalized to LPS-primed but NLRP3 activator-untreated controls.

Cellular fractionation and quantification of cytosolic mtDNA. Macrophages were first primed with LPS (200 ng ml⁻¹) followed by treatment with ATP or nigericin for 60 min or MSU for 3 h. Cellular fractionation was then performed using a mitochondrial isolation kit for cultured cells (89874, Thermo Fisher Scientific) according to manufacturer's instructions. Cytosolic mtDNA was measured as described¹⁰. In brief, DNA was isolated from 300 μ l of the cytosolic fractions (after normalization via cytosolic protein concentrations) of NLRP3 activator-treated BMDMs and mtDNA levels were quantified by qPCR as described above. For the measurement of ox-mtDNA, mtDNA was first purified using Allprep DNA/RNA mini kit (Qiagen) from the mitochondrial fraction of BMDMs as described above. The 8-OH-dG content of the mtDNA was then quantified using an 8-OHdG quantification kit (Cell Biolabs), as per the manufacturer's instruction.

Immunofluorescent staining and confocal microscopy. BMDMs were seeded at 0.2×10^6 cells per well in 8-well glass slides and rested overnight to allow proper attachment. To measure mtDNA replication, BMDMs were stimulated with LPS (200 ng ml⁻¹) for 3 or 6 h in the presence of 10 μ M EdU. To examine co-localization of newly synthesized mtDNA with inflammasome complexes, LPS-primed BMDMs were further treated with ATP and nigericin or transfected with poly (dA-dT) using Lipofectamine 2000. The cells were washed twice with sterile PBS and fixed in 2% paraformaldehyde (PFA) for 15 min followed by permeabilization with 0.1% Triton X-100 for 10 min. Endogenous peroxidases were blocked with 1% H₂O₂ in PBS for 30 min, followed by three washes with PBS. EdU staining was performed using a Click-iT EdU Microplate Assay Kit (Thermo Fisher Scientific). In brief, BMDMs were postfixed in EdU fixative for 5 min, and equal volumes of the EdU reaction cocktail, which was made immediately before use, were added to each chamber and incubated for 25 min. The cocktail was then removed, and BMDMs were washed three times in 1% blocking solution from the Click-iT EdU Microplate Assay Kit. The Oregon Green azide signal was then amplified with a TSA kit (Thermo Fisher Scientific). In brief, BMDMs were blocked in 1% bovine serum albumin (BSA) plus 5% normal goat serum in PBS for 60 min, followed by incubation with HRP-conjugated rabbit antibody against Oregon Green (from the EdU Microplate Assay Kit) diluted 1:300 in 1% BSA plus 5% goat serum in PBS overnight at 4 °C. For co-staining experiments, primary antibodies against ATP5B, ASC or 8-OH-dG (as described above) were included in the same solution with Oregon Green antibody. The next day, BMDMs were washed three times with PBS followed by incubation with Alexa-594 or -647 (from Life Technologies) secondary antibodies for 60 min. Then, BMDMs were stained with Alexa Fluor 488-labelled tyramide at 1:100 in amplification buffer plus 0.0015% H₂O₂ for 10 min, and washed three times with PBS. DAPI was used for nuclear counterstaining. Samples were imaged through a SP5 confocal microscope (Leica) 24 h after mounting.

CMCK2 reconstitution in *Irf1*^{-/-} macrophages. Wild-type and *Irf1*^{-/-} BMDMs were transduced with virus stocks containing either a wild-type (pFLRCmv-Yept-puro-mCMCK2) or a catalytically inactive (pFLRCmv-Yept-puro-mCMCK2-D330A) CMCK2-encoding lentivirus. Virus-containing supernatants were filtered through a 0.45- μ m-pore-size filter (Millipore) and supplemented with polybrene (8 μ g ml⁻¹) before adding to cells. Four days after viral transduction, successfully transduced BMDMs were selected by puromycin followed by analysis of mtDNA replication and NLRP3 inflammasome activation as described above.

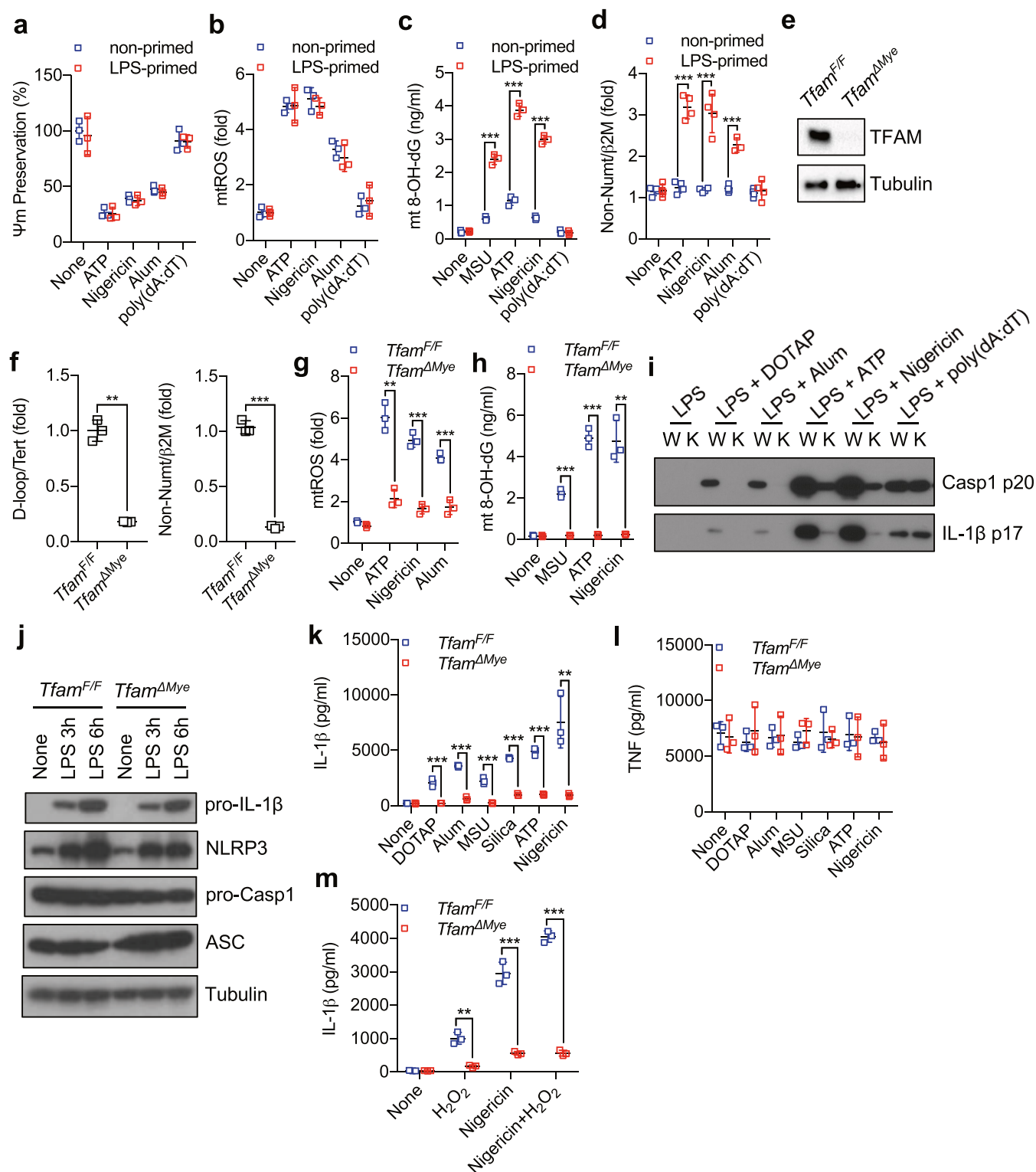
Septic shock and peritonitis models. Septic shock was induced by intraperitoneal injection of 8–12-week-old gender-matched wild-type and *Irf1*^{-/-} mice with LPS (*E. coli* O111:B4, Sigma-Aldrich) at 50 mg per kg body weight. Mouse survival was monitored every 6 h after injection for a total of 72 h. In separate experiments, mice were treated with the same dose of LPS and immune sera were collected 3 h post injection. Serum IL-1 β and TNF were measured by ELISA as described above. Peritonitis was induced by intraperitoneal injection of PBS or 1 mg alum (dissolved in 0.2 ml sterile PBS) into 8–12-week-old gender-matched wild-type and *Irf1*^{-/-} mice. Mice from each genotype were allocated randomly into PBS or alum treatment groups. After 12 h, mice were euthanized and the peritoneal cavities were washed with 6 ml cold sterile PBS. Neutrophils (CD11b⁺Ly6G⁺F4/80⁻) and monocytes (CD11b⁺Ly6C⁺Ly6G⁻) present in the peritoneal lavage fluid were quantified by flow cytometry. For blocking Fc-mediated interactions, mouse cells were pre-incubated with 0.5–1 μ g of purified anti-mouse CD16/CD32 per 100 μ l. Isolated cells were stained with labelled antibodies in PBS with 2% FCS and 2 mM EDTA or cell staining buffer (Biolegend). Dead cells were excluded based on staining with Live/Dead fixable dye (FVD-eFluor780, eBioscience). Absolute numbers of immune cell subtypes in the peritoneum were calculated by multiplying total peritoneal cell numbers by percentages of immune cell subtypes amongst total cells. Cells were analysed on a Beckman Coulter Cyan ADP flow cytometer. Data were analysed using FlowJo 10.2 software (Treestar). Flow cytometry gating strategy was shown in Supplementary Fig. 2. Antibodies specific for the following antigens were used: CD45 (m30-F11-V500); CD11b (mM1/70-eF450/eF660); MHCII (mM5/114.15.2-FITC/PE); Gr-1 (m1A8-Ly6G-PerCP-eF710); F4/80 (mBM8-PE/FITC); and Ly6C (mHK1.4-eF450) (from eBioscience and Biolegend).

Statistics. All data are mean \pm s.d. or mean \pm s.e.m. as indicated. Statistical analysis was performed using a two-tailed unpaired Student's *t*-test or log-rank test (for survival analysis). For all tests, $P < 0.05$ was considered statistically significant.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. All data are available from the corresponding author upon reasonable request.

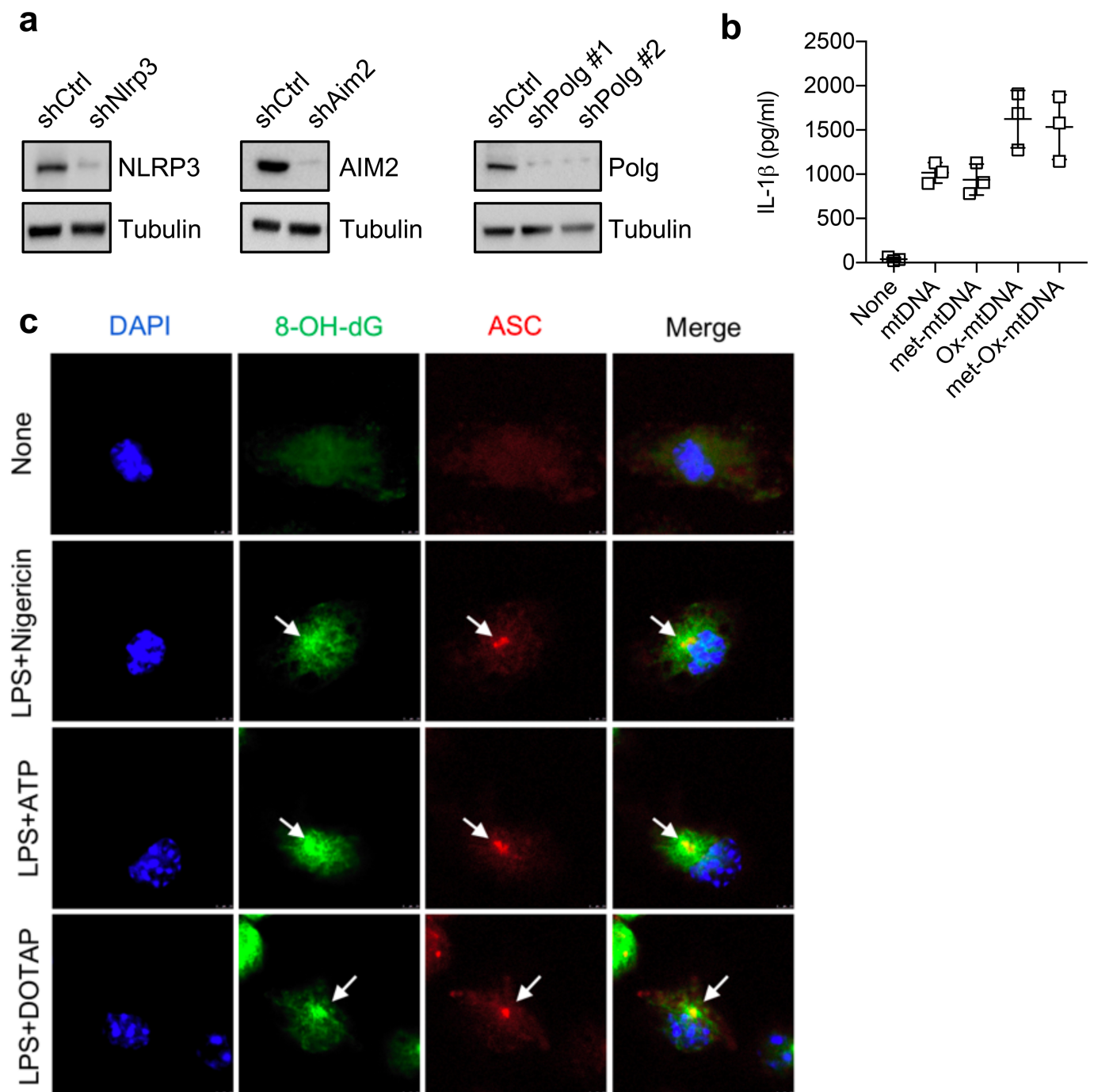
- Hudson, G. & Chinnery, P. F. Mitochondrial DNA polymerase-gamma and human disease. *Hum. Mol. Genet.* **15**, R244–R252 (2006).
- Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* **11**, 373–384 (2010).
- Ballana, E. & Esté, J. A. SAMHD1: at the crossroads of cell proliferation, immune responses, and virus restriction. *Trends Microbiol.* **23**, 680–692 (2015).
- Yamamoto, M. et al. Role of adaptor TRIF in the MyD88-independent Toll-like receptor signaling pathway. *Science* **301**, 640–643 (2003).
- Zhong, Z. et al. TRPM2 links oxidative stress to NLRP3 inflammasome activation. *Nat. Commun.* **4**, 1611 (2013).
- Hornung, V. et al. Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nat. Immunol.* **9**, 847–856 (2008).
- Malik, A. N., Czajka, A. & Cunningham, P. Accurate quantification of mouse mitochondrial DNA without co-amplification of nuclear mitochondrial insertion sequences. *Mitochondrion* **29**, 59–64 (2016).



Extended Data Fig. 1 | See next page for caption.

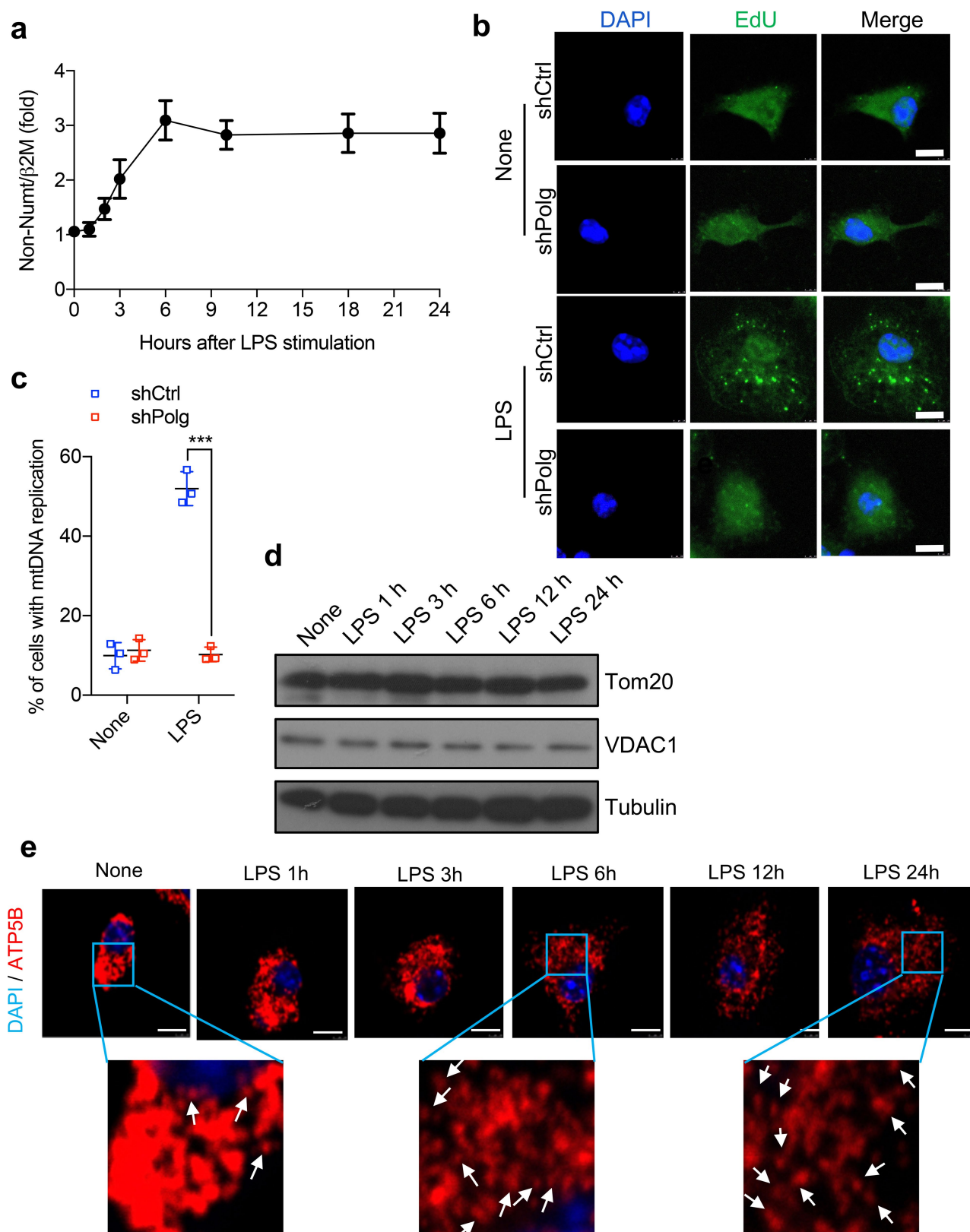
Extended Data Fig. 1 | TFAM is required for ox-mtDNA generation and NLRP3 inflammasome activation. **a**, Inflammasome activator-induced changes in mitochondrial membrane potential (Ψ_m) in non- or LPS-primed wild-type BMDMs were measured by TMRM fluorescence. Data are mean \pm s.d. ($n = 3$ biological replicates). **b**, Relative mtROS amounts were measured by MitoSOX fluorescence in non- or LPS-primed wild-type BMDMs after stimulation with different inflammasome activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **c**, Amounts of 8-OH-dG in mtDNA isolated from the mitochondrial fraction of non- or LPS-primed wild-type BMDMs that were treated with different inflammasome activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **d**, Cytosolic release of mtDNA, determined by qPCR with primers specific for mtDNA (non-NUMT) and nDNA (*B2m*), in non- or LPS-primed wild-type BMDMs after treatment with different inflammasome activators. Data are mean \pm s.d. ($n = 4$ biological replicates). **e**, Immunoblot analysis of TFAM in *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs. Results are typical of three independent experiments. **f**, Relative total mtDNA amounts in *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs determined by qPCR with primers specific for mtDNA (*D-loop*, non-NUMT) and nDNA (*Tert*, *B2m*). Data are mean \pm s.d. ($n = 3$ biological replicates). **g**, Relative mtROS amounts

were measured by MitoSOX fluorescence in LPS-primed *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs after stimulation with indicated NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **h**, Amounts of 8-OH-dG in mtDNA isolated from the mitochondrial fraction of LPS-primed *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs that were stimulated with various NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **i**, Immunoblot analysis of Casp1 p20 and mature IL-1 β (p17) in culture supernatants of *Tfam*^{fl/f} (W) and *Tfam* ^{Δ Mye} (K) BMDMs that were stimulated with LPS plus different inflammasome activators. Results are typical of three separate experiments. **j**, Immunoblot analysis of pro-IL-1 β , NLRP3, ASC and pro-Casp1 in the lysates of *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs before and after LPS priming. Results are typical of three separate experiments. **k**, **l**, Amounts of IL-1 β (**k**) and TNF (**l**) in culture supernatants of LPS-primed *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs that were stimulated with various NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **m**, Amounts of IL-1 β in culture supernatants of LPS-primed *Tfam*^{fl/f} and *Tfam* ^{Δ Mye} BMDMs that were stimulated with H₂O₂ in the absence and presence of nigericin. Data are mean \pm s.d. ($n = 3$ biological replicates). ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired *t*-test.



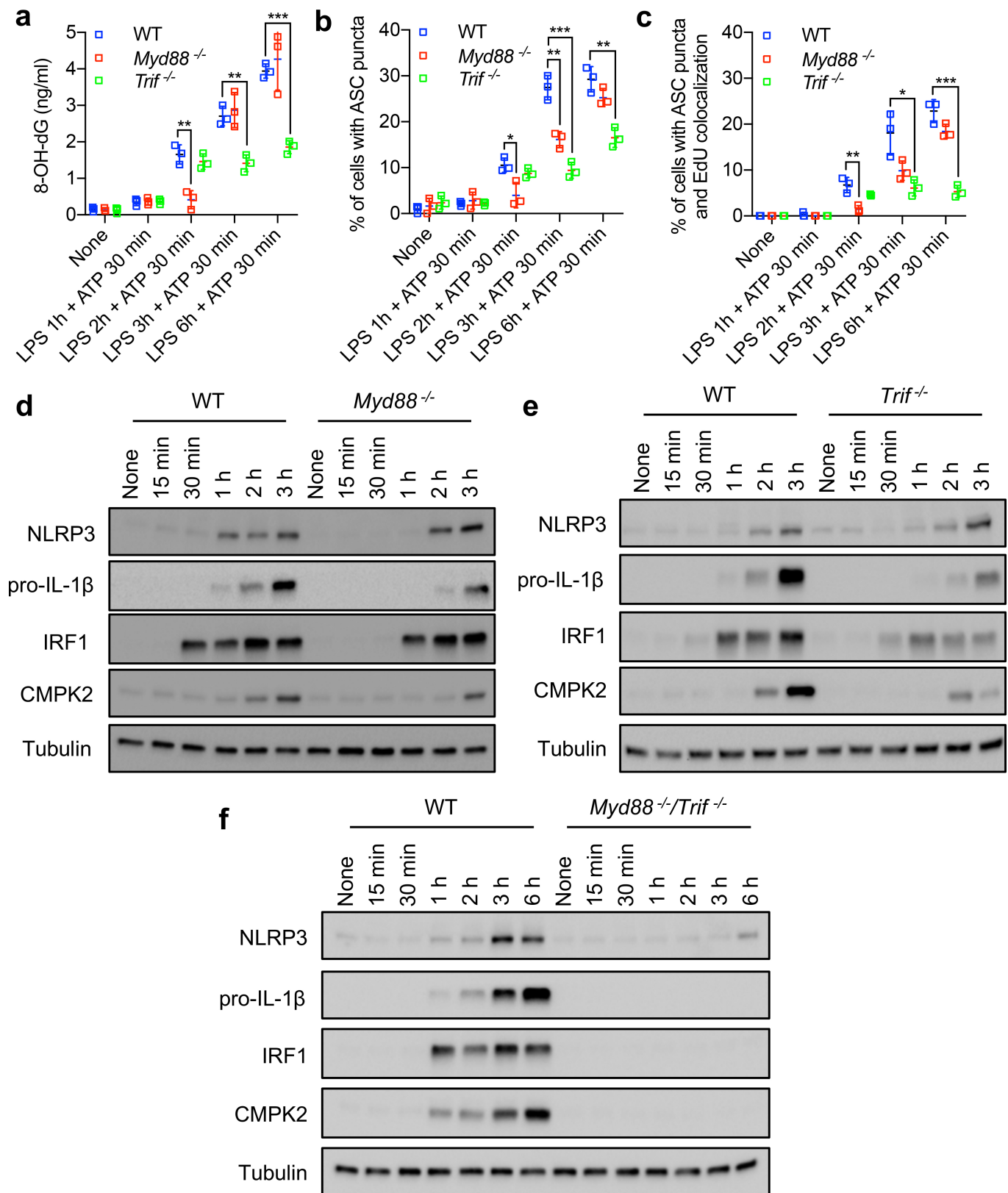
Extended Data Fig. 2 | ox-mtDNA activates the NLRP3 inflammasome. **a**, Immunoblot analysis of NLRP3, AIM2 and Pol γ in shCtrl- or specific shRNA-transduced BMDMs. Data are typical of three separate experiments. **b**, Amounts of IL-1 β in culture supernatants of LPS-primed *Tfam* Δ^{Myc} BMDMs that were transfected with mtDNA, methylated mtDNA, ox-mtDNA, and methylated ox-mtDNA. All mtDNAs were 90-bp

long and of an identical sequence. Data are mean \pm s.d. ($n = 3$ biological replicates). **c**, Representative fluorescent microscopy images of wild-type BMDMs that were co-stained for 8-OH-dG, ASC and DAPI before and after stimulation with LPS plus the indicated inflammasome activators. Results are typical of three independent experiments. Scale bars, 5 μ m.



Extended Data Fig. 3 | LPS induces mtDNA replication. **a**, Time-course analysis of total mtDNA amounts in wild-type BMDMs after LPS (200 ng ml^{-1}) stimulation. Data are mean \pm s.d. ($n = 3$ biological replicates). Results are typical of three separate experiments. **b**, Representative fluorescent microscopy images of EdU-labelled shCtrl- or shPolg-transduced wild-type BMDMs that were stimulated with or without LPS (200 ng ml^{-1}) for 6 h. Scale bars, $5 \mu\text{m}$. Results are typical of three separate experiments. **c**, Percentages of cells with mtDNA replication as determined in **b**. Data are mean \pm s.d. ($n = 3$ different microscopic fields per group;

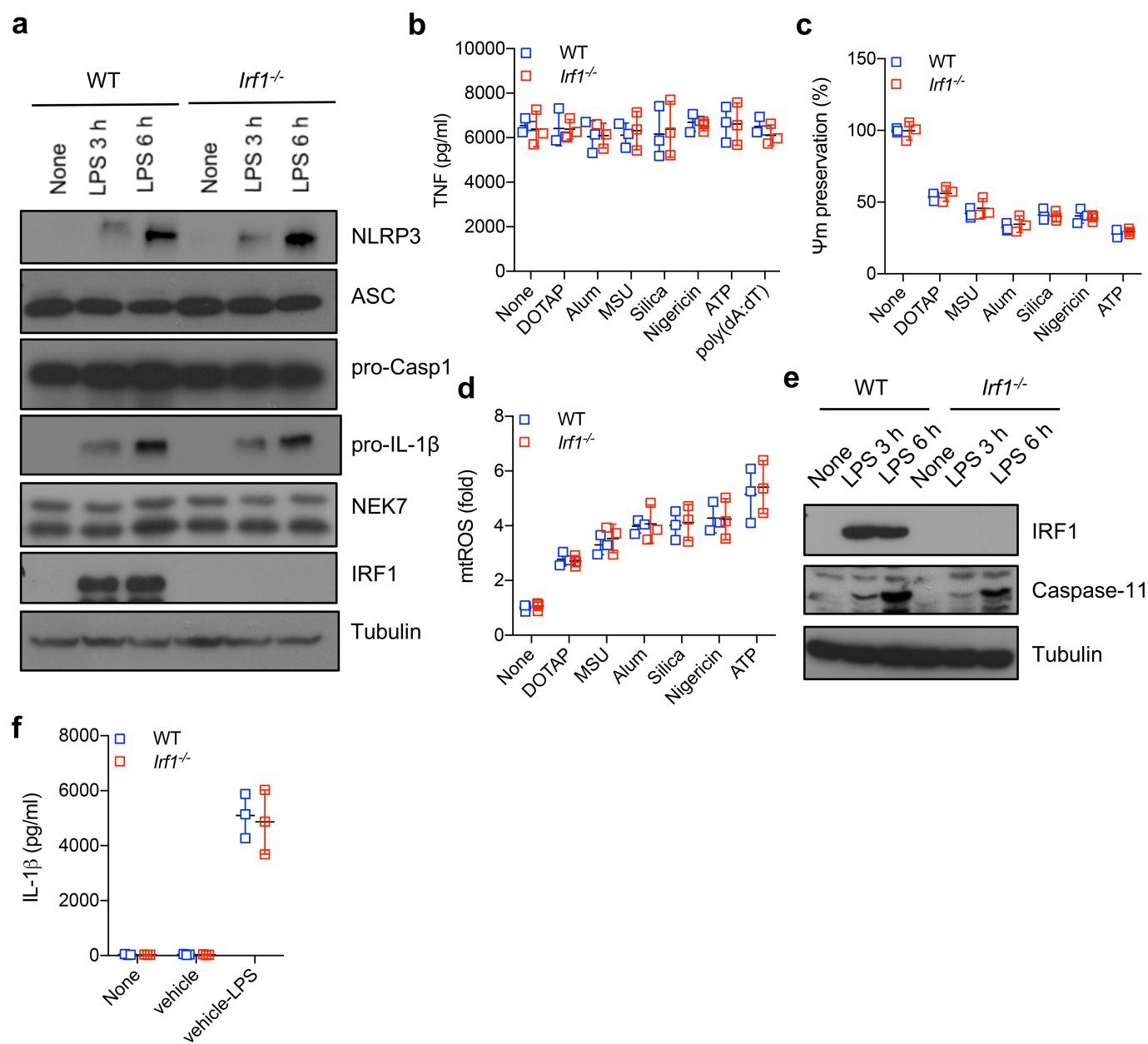
original magnification, $\times 40$). **d**, Immunoblot analysis of mitochondrial resident proteins, TOM20 and VDAC, in wild-type BMDMs before and after LPS stimulation. Results are typical of three separate experiments. **e**, Representative fluorescent microscopy images of the mitochondrial resident protein ATP5B and DAPI staining in wild-type BMDMs before and after LPS stimulation. Results are typical of three independent experiments. Arrows indicate fragmented mitochondria. *** $P < 0.001$, two-sided unpaired t -test.



Extended Data Fig. 4 | MyD88 and TRIF mediate LPS-induced IRF1 and CMPK2 expression and NLRP3 inflammasome activation.

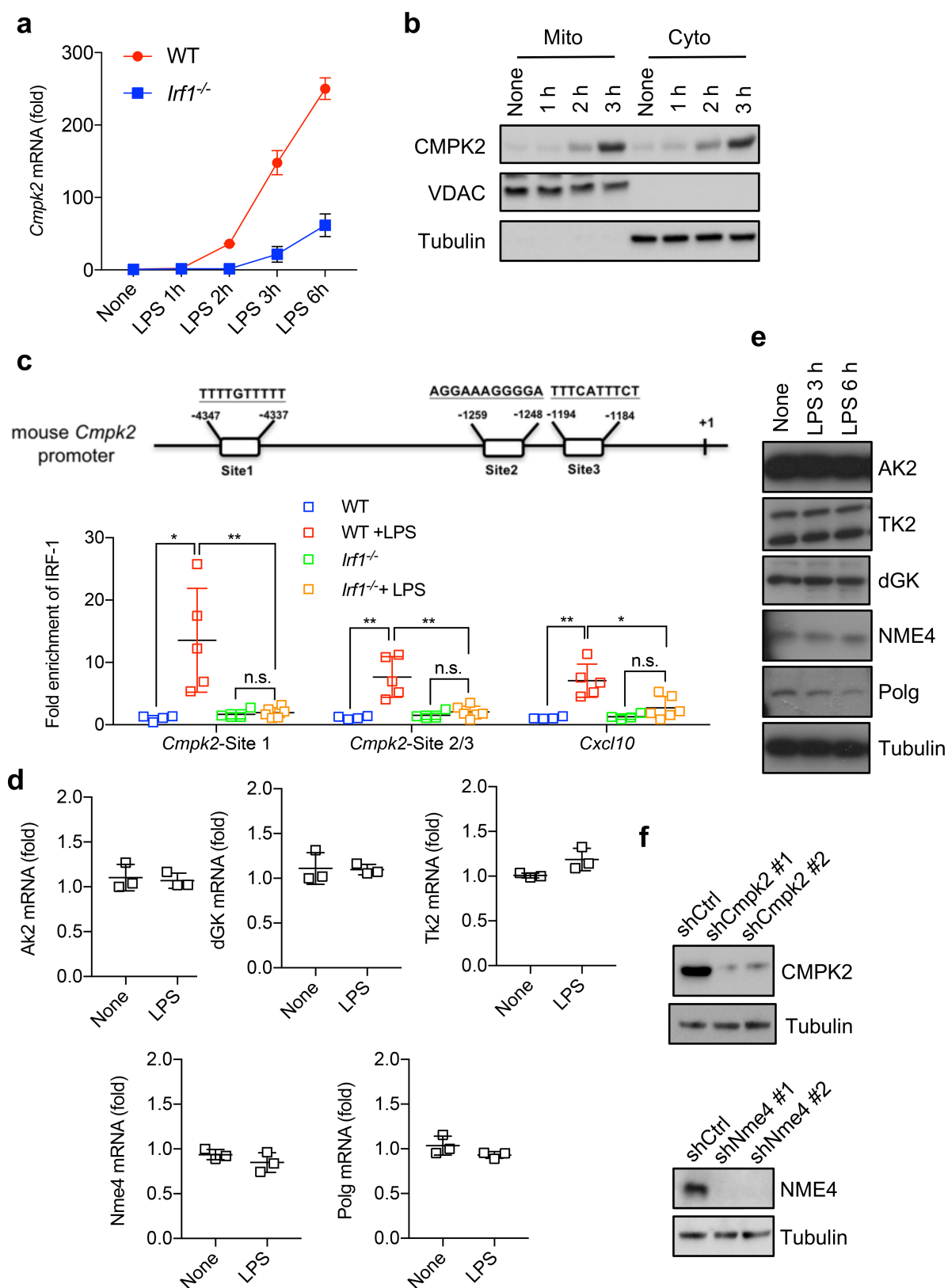
a, Amounts of 8-OH-dG in mtDNA isolated from mitochondrial fractions of wild-type, *Myd88*^{-/-} and *Trif*^{-/-} BMDMs that were primed with LPS for different durations followed by stimulation with ATP. Data are mean \pm s.d. ($n = 3$ biological replicates per time point). **b**, **c**, Quantification of fluorescent microscopy images of ASC puncta (**b**) and ASC puncta

positive for EdU (**c**) in wild-type, *Myd88*^{-/-} and *Trif*^{-/-} BMDMs that were primed with LPS for different durations, followed by ATP stimulation. Data are mean \pm s.d. ($n = 3$ different microscopic fields per group; original magnification, $\times 40$). **d–f**, Immunoblot analysis of NLRP3, pro-IL-1 β , IRF1 and CMPK2 in wild-type, *Myd88*^{-/-} and *Trif*^{-/-} BMDMs before and after LPS stimulation. Data are typical of three separate experiments. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired *t*-test.



Extended Data Fig. 5 | Effects of IRF1 on priming, mitochondrial damage and non-canonical inflammasome activation. **a**, Immunoblot analysis of NLRP3, ASC, pro-caspase-1, pro-IL-1β, NEK7 and IRF1 in the lysates of wild-type and *Irf1*^{-/-} BMDMs before and after LPS stimulation. Data are typical of three separate experiments. **b**, Amounts of TNF in culture supernatants of LPS-primed wild-type and *Irf1*^{-/-} BMDMs that were stimulated with the indicated inflammasome activators. Data are mean ± s.d. (*n* = 3 biological replicates). **c**, Percentages of ψ_m preservation were measured by TMRM fluorescence in LPS-primed wild-type and *Irf1*^{-/-} BMDMs that were stimulated with indicated NLRP3 activators.

Data are mean ± s.d. (*n* = 3 biological replicates). **d**, Relative amounts of mtROS were measured by MitoSOX fluorescence in LPS-primed wild-type and *Irf1*^{-/-} BMDMs that were stimulated with indicated NLRP3 activators. Data are mean ± s.d. (*n* = 3 biological replicates). **e**, Immunoblot analysis of caspase-11 and IRF1 in lysates from wild-type and *Irf1*^{-/-} BMDMs before and after LPS stimulation. Results are typical of three independent experiments. **f**, Amounts of IL-1β in culture supernatants of LPS-primed wild-type and *Irf1*^{-/-} BMDMs that were further transfected with FuGENE-complexed LPS. Data are mean ± s.d. (*n* = 3 biological replicates).

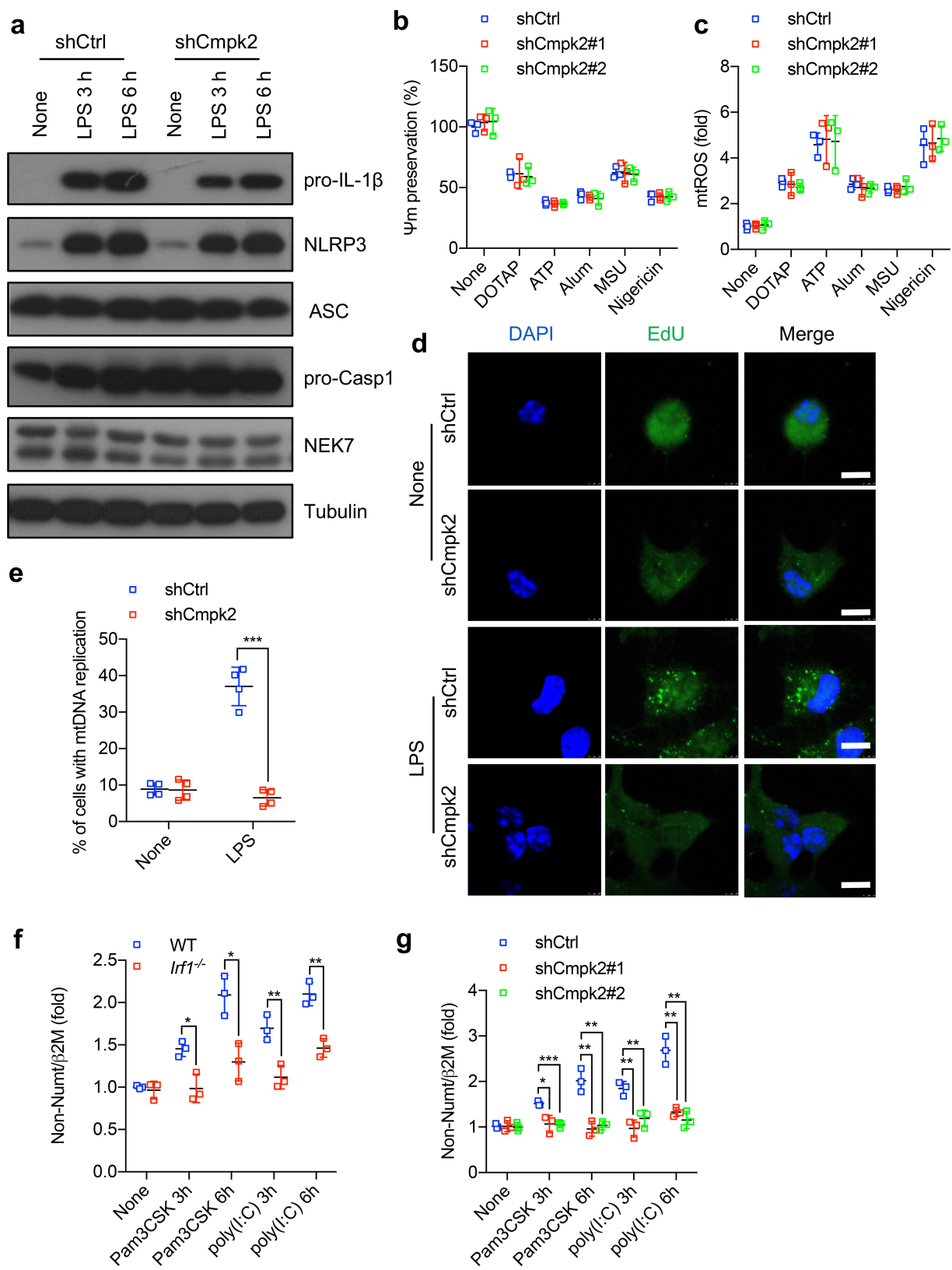


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | IRF1 mediates LPS-induced CMPK2 expression.

a, Relative amounts of *Cmpk2* mRNA in wild-type and *Irf1*^{-/-} BMDMs before and after LPS stimulation. Data are mean ± s.d. (*n* = 3 biological replicates per time point). **b**, Immunoblot analysis of CMPK2, VDAC and tubulin in mitochondrial and cytosolic fractions of wild-type BMDMs after LPS stimulation. Results are typical of three separate experiments. **c**, Chromatin immunoprecipitation analysis of IRF1 recruitment to the *Cmpk2* promoter. Data are mean ± s.d. (*n* = 4 biological replicates for wild-type and *Irf1*^{-/-} groups; *n* = 5 and 6 biological replicates for wild-type + LPS and *Irf1*^{-/-} + LPS groups, respectively). *Cxcl10*, a known

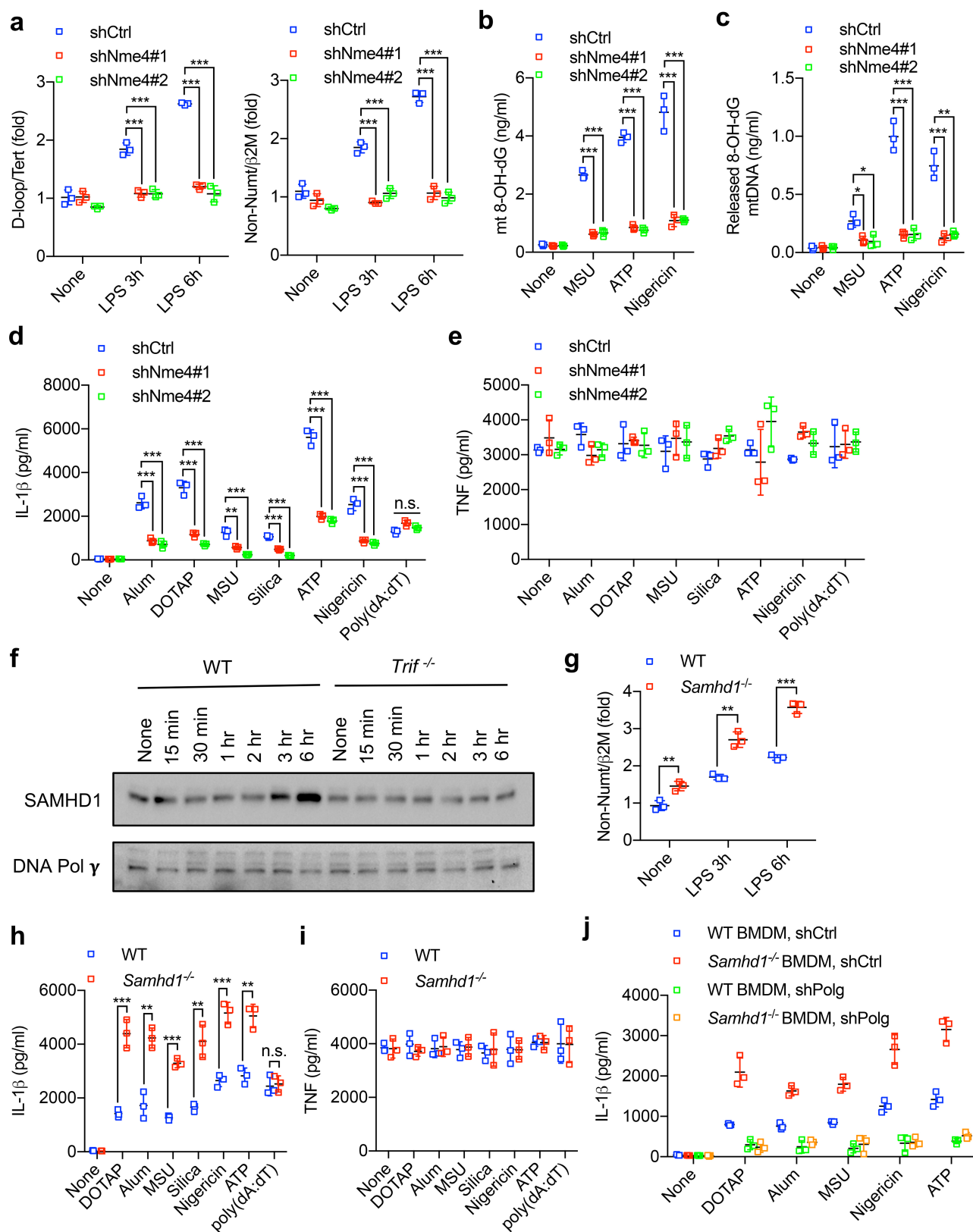
IRF1-target gene, was included as a positive control. **d**, Relative mRNA amounts of *dGK* (also known as *Dguok*), *Tk2*, *Ak2*, *Nme4* and *Polg* in wild-type BMDMs before and after 6 h LPS stimulation. Data are mean ± s.d. (*n* = 3 biological replicates). **e**, Immunoblot analysis of the enzymes encoded by the genes in **d** in the lysates of wild-type BMDMs before and after LPS stimulation. Results are typical of three independent experiments. **f**, Immunoblot analysis of CMPK2 and NME4 in shCtrl- or specific shRNA-transduced BMDMs. Data are typical of three independent experiments. **P* < 0.05; ***P* < 0.01; two-sided unpaired *t*-test.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | CMPK2 deficiency does not affect inflammasome subunit expression nor NLRP3 activator-induced mitochondrial damage. **a**, Immunoblot analysis of pro-IL-1 β , NLRP3, ASC, pro-Casp1 and NEK7 in the lysates of wild-type (shCtrl) and CMPK2-deficient (shCmpk2) BMDMs before and after LPS priming. Results are typical of three separate experiments. **b**, NLRP3 activator-induced changes in Ψ_m in LPS-primed shCtrl and shCmpk2 BMDMs were measured by TMRM fluorescence. Data are mean \pm s.d. ($n = 3$ biological replicates). **c**, Relative amounts of mtROS measured by MitoSOX fluorescence in LPS-primed shCtrl and shCmpk2 BMDMs after stimulation with the indicated NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **d**, Representative fluorescent microscopy

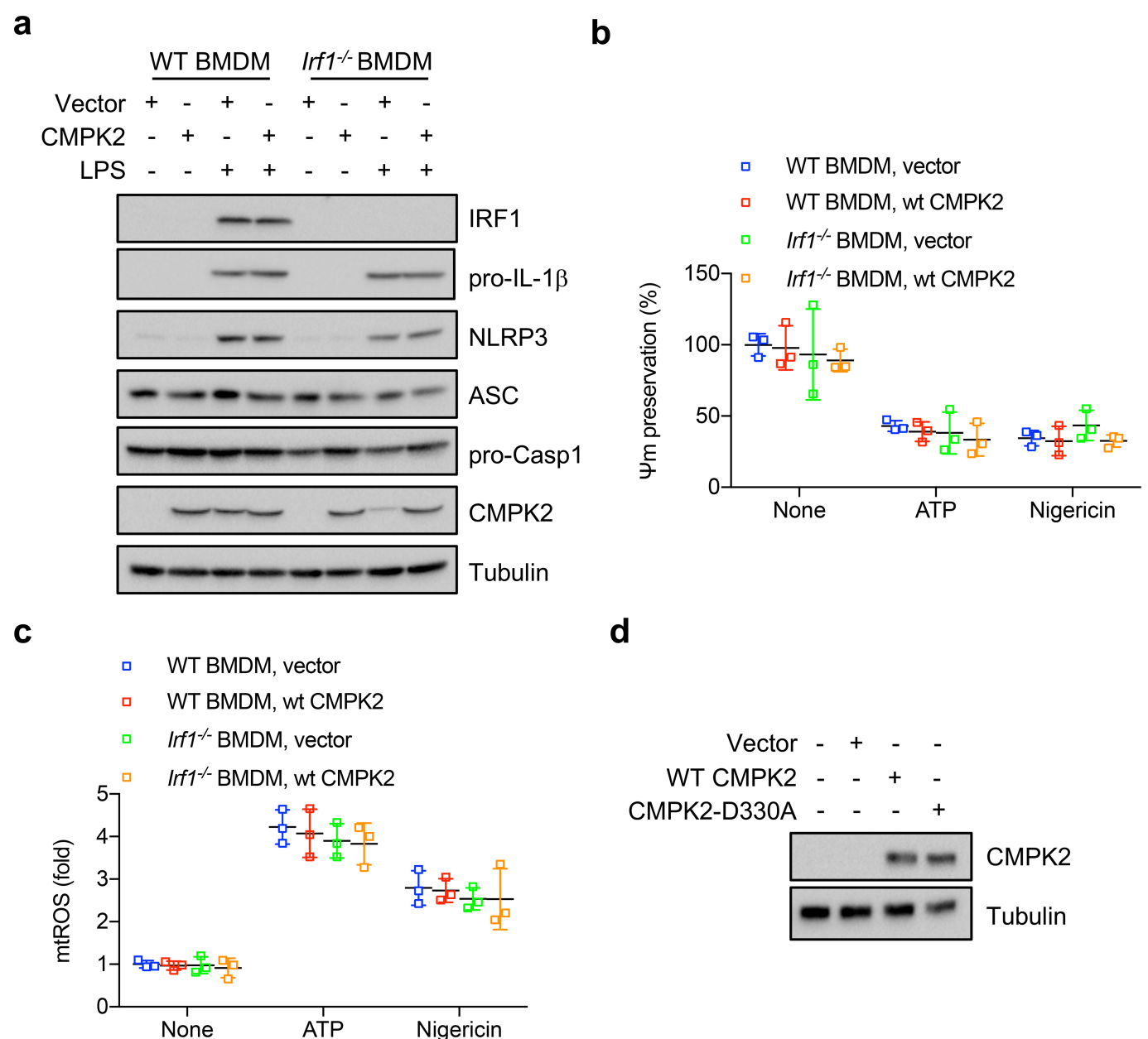
images of EdU-labelled wild-type BMDMs transduced with either shCtrl- or shCmpk2-encoding lentiviruses that were stimulated with or without LPS (200 ng ml^{-1}) for 6 h. Scale bars, $5 \mu\text{m}$. **e**, Percentages of cells with mtDNA replication as determined in **d**. Data are mean \pm s.d. ($n = 3$ different microscopic fields per group; original magnification, $\times 40$). **f**, Relative amounts of total mtDNA in wild-type and *Irf1* $^{-/-}$ BMDMs before and after treatments with the indicated TLR agonists. Data are mean \pm s.d. ($n = 3$ biological replicates). **g**, Relative amounts of total mtDNA in shCtrl- or shCmpk2-encoding lentivirus-transduced wild-type BMDMs before and after treatments with the indicated TLR agonists. Data are mean \pm s.d. ($n = 3$ biological replicates). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired t -test.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | dNTP availability controls LPS-induced mtDNA synthesis and NLRP3 inflammasome activation. **a**, Relative total mtDNA amounts in shCtrl and shNme4 BMDMs before and after LPS priming. Data are mean \pm s.d. ($n = 3$ biological replicates). **b**, Amounts of 8-OH-dG in mtDNA isolated from the mitochondrial fraction of LPS-primed shCtrl and shNme4 BMDMs that were stimulated with various NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **c**, Amounts of 8-OH-dG in cytosolic mtDNA from LPS-primed shCtrl and shNme4 BMDMs that were stimulated with various NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **d**, **e**, Amounts of IL-1 β (**d**) and TNF (**e**) in supernatants of LPS-primed shCtrl and shNme4 BMDMs that were stimulated with various inflammasome activators. Data are mean \pm s.d. ($n = 3$ biological replicates). **f**, Immunoblot analysis of

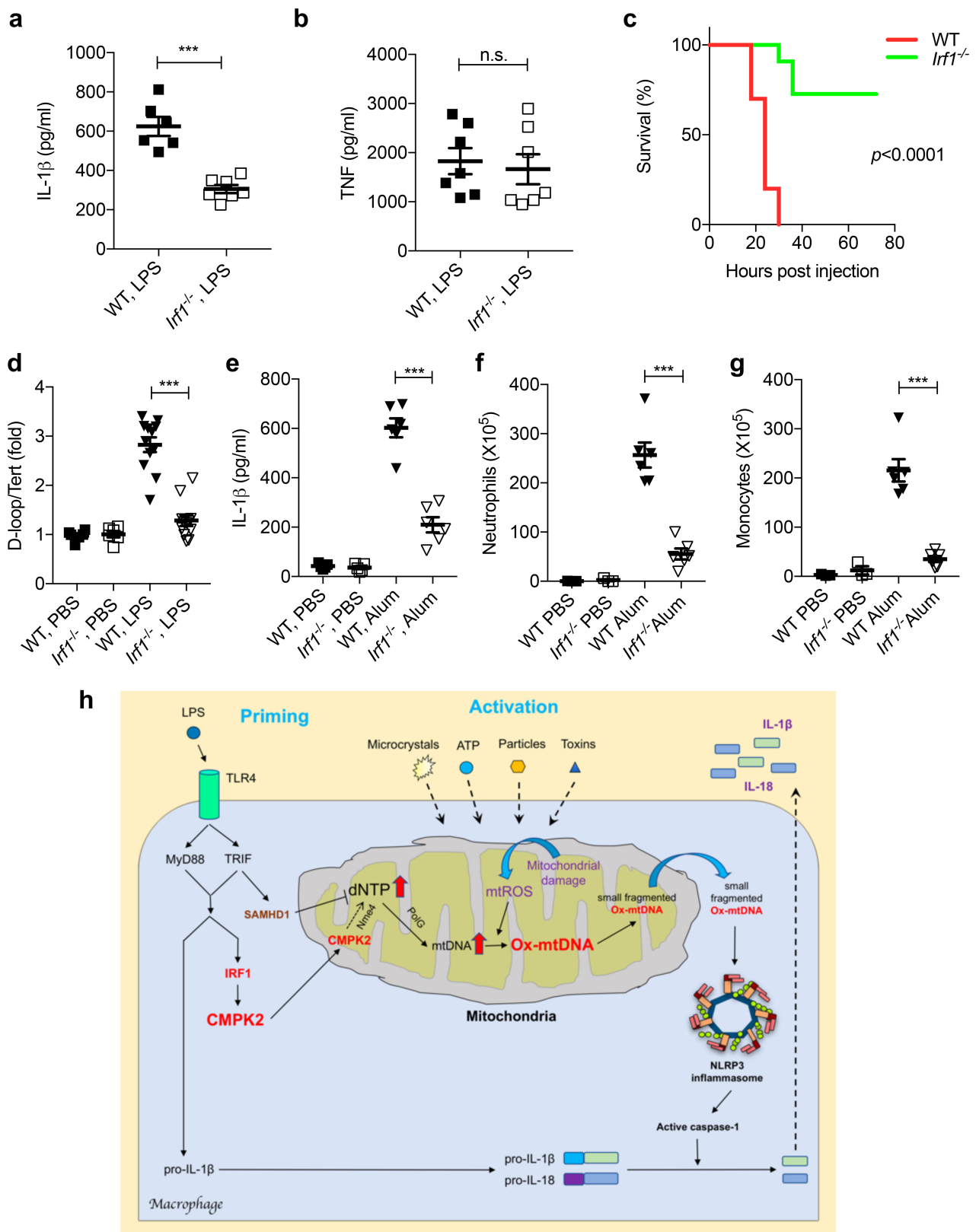
SAMHD1 and Pol γ in wild-type and *Trif*^{-/-} BMDMs that were stimulated with LPS for different durations as indicated. Results are typical of three independent experiments. **g**, Relative total mtDNA amounts in wild-type and *Samhd1*^{-/-} BMDMs before and after LPS stimulation. Data are mean \pm s.d. ($n = 3$ biological replicates). **h**, **i**, Amounts of IL-1 β (**h**) and TNF (**i**) in the culture supernatants of LPS-primed wild-type and *Samhd1*^{-/-} BMDMs that were stimulated with inflammasome activators as indicated. Data are mean \pm s.d. ($n = 3$ biological replicates). **j**, Amounts of NLRP3 activator-induced IL-1 β in culture supernatants of LPS-primed wild-type and *Samhd1*^{-/-} BMDMs with or without *Polg* expression. Data are mean \pm s.d. ($n = 3$ biological replicates). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; two-sided unpaired t -test.



Extended Data Fig. 9 | CMPK2 expression restores NLRP3 inflammasome activation in IRF1-deficient macrophages.

a, Immunoblot analysis of IRF1, CMPK2, pro-IL-1 β , NLRP3, ASC and pro-Casp1 in lysates of wild-type and *Irf1*^{-/-} BMDMs before and after transduction with a wild-type CMPK2-encoding lentivirus. Results are typical of three independent experiments. **b**, NLRP3 activator-induced changes in Ψ_m in LPS-primed CMPK2-transduced wild-type and *Irf1*^{-/-} BMDMs were measured by TMRM fluorescence. Data are mean \pm s.d. ($n = 3$ biological replicates) and analysed by two-sided unpaired *t*-test

(not significant). **c**, Relative amounts of mtROS measured by MitoSOX fluorescence in LPS-primed control (vector)- or CMPK2-transduced wild-type and *Irf1*^{-/-} BMDMs before and after stimulation with NLRP3 activators. Data are mean \pm s.d. ($n = 3$ biological replicates) and analysed by two-sided unpaired *t*-test (not significant). **d**, Immunoblot analysis of CMPK2 in *Irf1*^{-/-} BMDMs that were transduced with wild-type or mutant (CMPK2(D330A)) CMPK2-encoding lentiviruses. Results are typical of three separate experiments.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | IRF1 is required for in vivo mtDNA replication and NLRP3 inflammasome activation. **a, b**, 12-week-old wild-type or *Irf1*^{-/-} mice were injected intraperitoneally with LPS (50 mg per kg of body weight) and their sera were collected 3 h later and analysed by ELISA for IL-1 β (**a**) and TNF (**b**). Results are mean \pm s.d. ($n = 6$ and 7 for WT and *Irf1*^{-/-} mice, respectively). **c**, Survival of wild-type or *Irf1*^{-/-} mice that were injected intraperitoneally with LPS (50 mg per kg body weight; $n = 10$ and 11 for WT and *Irf1*^{-/-} mice, respectively). **d**, Relative amounts of total mtDNA in peritoneal infiltrates of wild-type or *Irf1*^{-/-} mice before and after LPS (50 mg per kg body weight) injection. Data are mean \pm s.d. ($n = 6$ in PBS-treated groups; $n = 12$ in LPS-treated groups). **e**, Peritoneal IL-1 β in wild-type or *Irf1*^{-/-} mice 4 h after intraperitoneal injection

of alum (1 mg) or PBS. Data are mean \pm s.d. ($n = 5$ in PBS-treated groups; $n = 6$ in alum-treated groups). **f, g**, Alum-induced peritoneal infiltration of neutrophils (CD11b⁺Ly6G⁺F4/80⁻) (**f**) and monocytes (CD11b⁺Ly6C⁺Ly6G⁻) (**g**) in wild-type and *Irf1*^{-/-} mice 12 h after alum (1 mg) or PBS injection. Data are mean \pm s.e.m. ($n = 3$ for PBS-treated groups and $n = 6$ for alum-treated groups). *** $P < 0.001$; two-sided unpaired *t*-test (**a, b, d–g**) and log-rank test (**c**). **h**, A working model to illustrate how TLR-mediated priming controls mtDNA replication and NLRP3 inflammasome activation. Whereas IRF1 acts positively to induce the transcription of CMPK2, which supplies rate-limiting dCDP for mtDNA synthesis, TRIF-dependent signalling also acts negatively to limit dNTP supply through the induction of SAMHD1.

Topological band engineering of graphene nanoribbons

Daniel J. Rizzo^{1,5}, Gregory Veber^{2,5}, Ting Cao^{1,3,5}, Christopher Bronner¹, Ting Chen¹, Fangzhou Zhao¹, Henry Rodriguez¹, Steven G. Louie^{1,3*}, Michael F. Crommie^{1,3,4*} & Felix R. Fischer^{2,3,4*}

Topological insulators are an emerging class of materials that host highly robust in-gap surface or interface states while maintaining an insulating bulk^{1,2}. Most advances in this field have focused on topological insulators and related topological crystalline insulators³ in two dimensions^{4–6} and three dimensions^{7–10}, but more recent theoretical work has predicted the existence of one-dimensional symmetry-protected topological phases in graphene nanoribbons (GNRs)¹¹. The topological phase of these laterally confined, semiconducting strips of graphene is determined by their width, edge shape and terminating crystallographic unit cell and is characterized by a \mathbb{Z}_2 invariant¹² (that is, an index of either 0 or 1, indicating two topological classes—similar to quasi-one-dimensional solitonic systems^{13–16}). Interfaces between topologically distinct GNRs characterized by different values of \mathbb{Z}_2 are predicted to support half-filled, in-gap localized electronic states that could, in principle, be used as a tool for material engineering¹¹. Here we present the rational design and experimental realization of a topologically engineered GNR superlattice that hosts a one-dimensional array of such states, thus generating otherwise inaccessible electronic structures. This strategy also enables new end states to be engineered directly into the termini of the one-dimensional GNR superlattice. Atomically precise topological GNR superlattices were synthesized from molecular precursors on a gold surface, Au(111), under ultrahigh-vacuum conditions and characterized by low-temperature scanning tunnelling microscopy and spectroscopy. Our experimental results and first-principles calculations reveal that the frontier band structure (the bands bracketing filled and empty states) of these GNR superlattices is defined purely by the coupling between adjacent topological interface states. This manifestation of non-trivial one-dimensional topological phases presents a route to band engineering in one-dimensional materials based on precise control of their electronic topology, and is a promising platform for studies of one-dimensional quantum spin physics.

GNRs represent a promising scaffold in the exploration of topological phases because graphene becomes semiconducting when confined laterally with certain edge structures^{17,18}. Recent advancements in the rational bottom-up synthesis of GNRs have provided atomically precise control over almost all structural parameters through the rational design and self-assembly of small-molecule precursors¹⁹. This has allowed exploration of GNR energy-gap versus width relations^{17,18} and bandgap engineering via dopant-mediated shifts in electron affinity^{20–25}. Fusion of different types of GNR precursors along the longitudinal axis has led to the design and synthesis of type I and type II heterojunctions where GNR electronic structure changes continuously from one GNR type to another as the heterojunction interface is crossed^{21,26}.

Topological concepts, on the other hand, provide a different strategy in the design of bottom-up GNR electronic structure. We exploit the nontrivial topological phases determined by a \mathbb{Z}_2 invariant associated

with the width, edge shape and the termination of the GNR¹¹. Robust half-occupied interface states are predicted to occur at the heterojunctions between topologically trivial and nontrivial GNR segments (that is, where the value of \mathbb{Z}_2 changes across an interface, as shown in Fig. 1a). These interface states, if aligned periodically in a superlattice, enable a hierarchy of topological quantum engineering because they are defined locally by topological phase discontinuities, whereas the superlattice's global electronic structure reflects the hybridization between them. The end properties of such a GNR superlattice are determined by the topology of the overall superlattice electronic structure. This scheme provides new strategies for modifying GNR bandgaps and even potentially inducing completely new GNR behaviours such as metallicity and magnetism out of individually semiconducting structural components.

Our strategy for topologically engineering new bottom-up GNR behaviour relies on the synthesis of atomically precise superlattices that are comprised of alternating topologically trivial 7-armchair GNR (AGNR) segments ($\mathbb{Z}_2 = 0$) and topologically nontrivial 9-AGNR segments ($\mathbb{Z}_2 = 1$) along the longitudinal GNR axis, thus leading to a one-dimensional array of interface states¹¹. If the coupling between these states is expressed as hopping amplitudes t_1 (for hopping across a 9-AGNR segment) and t_2 (for hopping across a 7-AGNR segment), then the band dispersion arising from the coupled topological interface states can be expressed in the standard two-band tight-binding form¹⁶

$$E_{\pm}(k) = \pm \sqrt{t_1^2 + t_2^2 + 2t_1t_2 \cos(k)} \quad (1)$$

This leads to a tunable energy gap ($E_g = 2 ||t_1| - |t_2||$) and bandwidth ($W = |t_1| + |t_2| - E_g/2$) for the new bands that arise from purely topological considerations. The end properties of this superlattice, however, are not solely determined by t_1 and t_2 , but must take into account the Zak or Berry phase of all the occupied π -electron bands¹¹. By carefully controlling the atomic structure of the 7/9-AGNR superlattice termini, we ensure that the resulting global topological phase of the entire 7/9-AGNR superlattice is nontrivial. This mandates the existence of a series of end states in different energy gaps of this hierarchically engineered one-dimensional topological system.

The key to creating well defined, periodic topological interface states is the design of molecular precursors that selectively link crystallographic unit cells of 7-AGNRs and 9-AGNRs into segments that have different topological phases. This is achieved by controlling the molecular structure of the 7/9-AGNR interface through careful design of building block **1** (Fig. 1b) (even small changes in the alignment of the interface structure can alter the topological phase of the constituent GNR segments¹¹). The structural asymmetry between the two distinct reaction interfaces in **1** (a zigzag edge on the side of the 7-AGNR and an armchair edge on the side of the 9-AGNR) leads to a sterically enforced highly selective head-to-head or tail-to-tail polymerization during the on-surface synthesis (Fig. 1c, Extended Data Fig. 1). Thermally induced

¹Department of Physics, University of California, Berkeley, CA, USA. ²Department of Chemistry, University of California, Berkeley, CA, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁴Kavli Energy NanoSciences Institute at the University of California Berkeley and the Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁵These authors contributed equally: Daniel J. Rizzo, Gregory Veber, Ting Cao. *e-mail: sglouie@berkeley.edu; crommie@berkeley.edu; ffischer@berkeley.edu

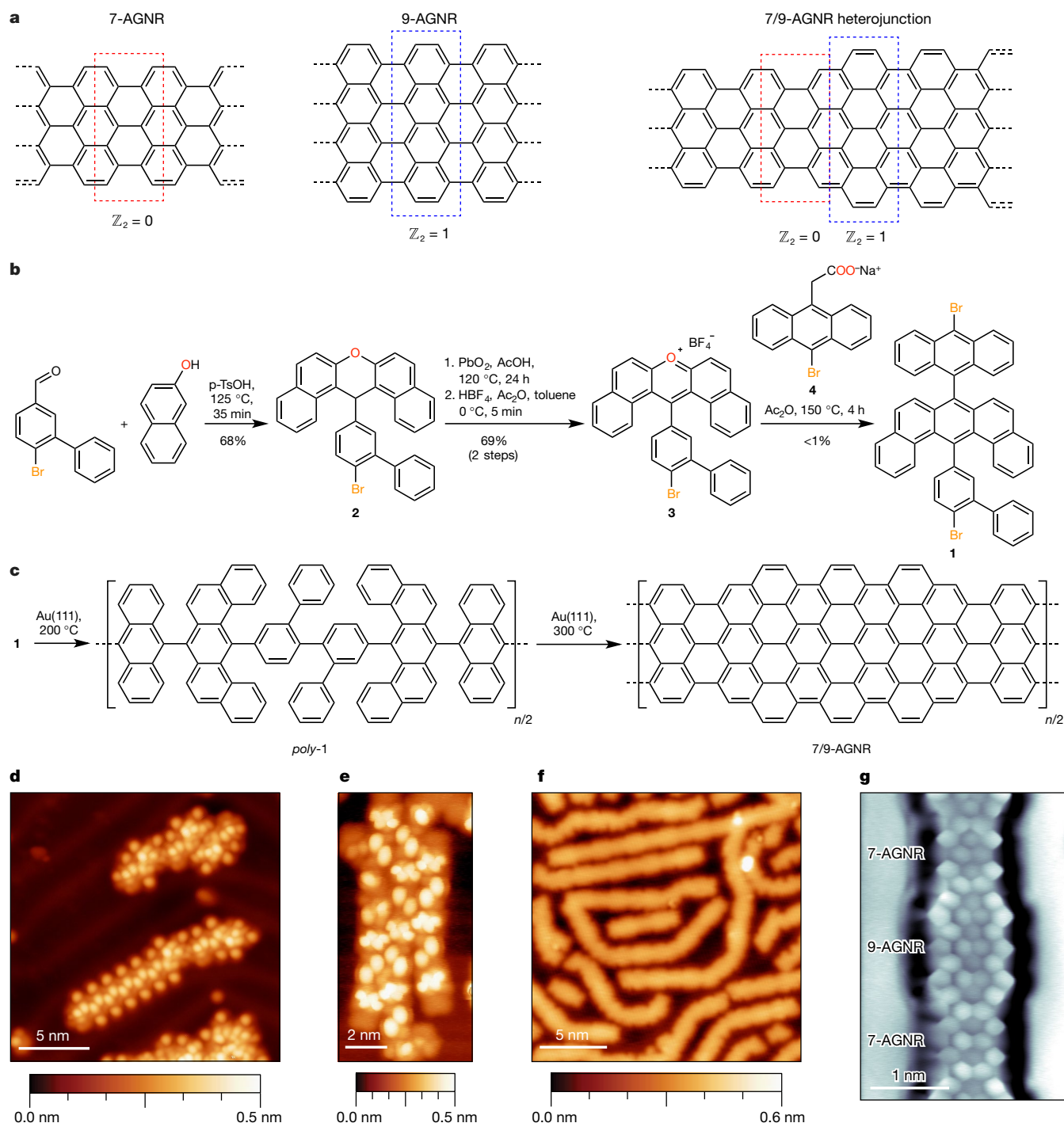


Fig. 1 | Bottom-up synthesis of 7/9-AGNR superlattice on Au(111).

a, Schematic representation of the \mathbb{Z}_2 invariant associated with particular terminations of 7- and 9-AGNRs and at the interface of a 7/9-AGNR heterojunction. **b**, Synthesis of molecular precursor **1** (p-TsOH is p-toluenesulfonic acid, AcOH is acetic acid and Ac₂O is acetic anhydride). **c**, Schematic representation of the stepwise thermally induced on-surface growth of a 7/9-AGNR superlattice from molecular precursor **1**. *Poly-1* is formed upon annealing at 200 °C and full cyclization occurs at 300 °C. Activated precursors polymerize in a head-to-head orientation owing to

the distinctive steric constraints of the two active sites. **d**, STM topography of precursor **1** as deposited on Au(111) (sample bias $V_s = 1.00$ V, tunnelling current $I_t = 30$ pA). **e**, Polymer island on Au(111) after annealing at 200 °C ($V_s = 1.00$ V, $I_t = 30$ pA). **f**, Fully cyclized 7/9-AGNR superlattice on Au(111) after annealing at 300 °C ($V_s = 0.20$ V, $I_t = 30$ pA). **g**, A bond-resolved STM image of 7/9-AGNR superlattice shows the bond-resolved structure of the heterojunction interface ($V_s = 0.02$ V, $I_t = 80$ pA, bias modulation frequency $f = 581$ Hz, bias modulation amplitude $V_{ac} = 12$ mV). All STM data were obtained at $T = 4$ K.

cyclodehydrogenation of the resulting 7/9-polymer intermediates yields the topologically trivial/nontrivial superlattice of 7/9-AGNRs (Fig. 1c).

The synthesis of **1** is depicted in Fig. 1b. Condensation of 6-bromo-(1,1'-biphenyl)-3-carbaldehyde with 2-naphthol yields xanthene **2** in 68% yield. Benzylic oxidation with lead(IV) oxide followed by dehydration of the intermediate xanthanol with

tetrafluoroboric acid gives the pyrylium salt **3** in 69% over two steps. The molecular precursor **1** can be obtained as a minor product (<10% crude mixture, major product is the xanthanol) from the condensation of **3** with the sodium salt of 2-(10-bromoanthracene-9-yl)acetic acid. Analytically pure samples suitable for ultrahigh-vacuum deposition were obtained through multiple precipitations and recrystallizations

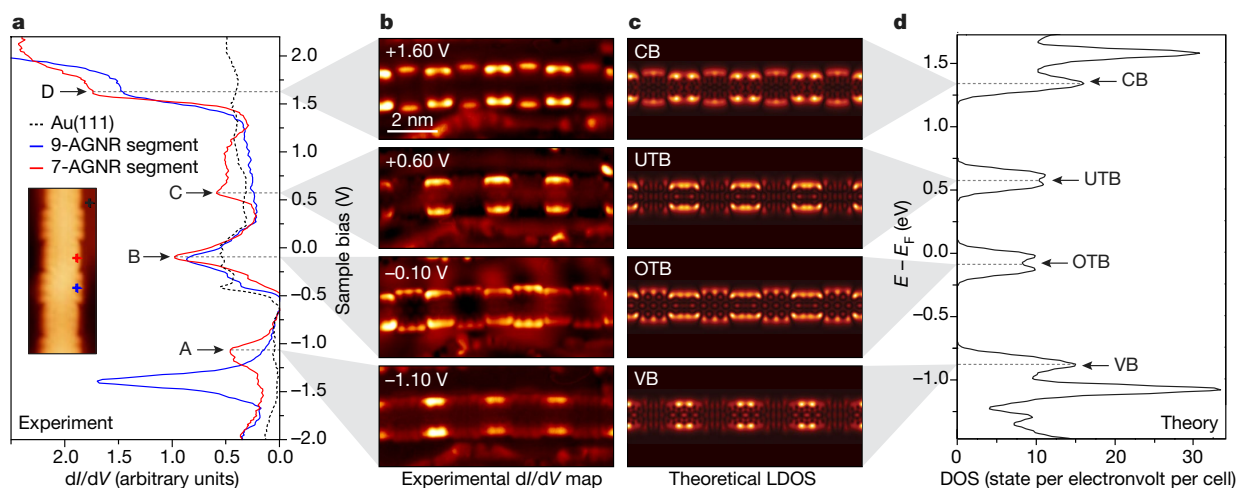


Fig. 2 | Electronic structure of 7/9-AGNR superlattice. **a**, Inset, STM topography of 7/9-AGNR superlattice ($V_s = -0.10$ V, $I_t = 80$ pA). The red (blue) curve shows dI/dV point spectroscopy data collected on the 7-AGNR (9-AGNR) segment (tip position marked by plus sign). The dotted black curve shows a spectrum collected on bare Au(111). The spectroscopy parameters are $V_{ac} = 10$ mV and $f = 581$ Hz. **b**, Constant-current dI/dV maps obtained at voltages corresponding to peaks A, B,

C and D marked by arrows in **a** ($I_t = 100$ pA, $V_{ac} = 10$ mV, $f = 581$ Hz).

c, LDOS maps calculated via DFT at the energies of the 7/9-AGNR superlattice VB, OTB, UTB and CB marked by arrows in **d** (obtained 4 Å above the GNR plane). **d**, Calculated DOS obtained via DFT for the 7/9-AGNR superlattice (0.05 eV Gaussian broadening used). All STM data shown here were obtained at $T = 4$ K on a 7/9-AGNR consisting of 12 fused monomer units (the total GNR length was 15.6 nm).

from EtOH/ CHCl_3 , 7/9-AGNR superlattices were grown on Au(111) by sublimation of **1** onto a clean Au(111) single crystal under ultrahigh-vacuum conditions. Figure 1d shows a scanning tunnelling microscopy (STM) image of a sub-monolayer coverage of molecular precursor **1** on Au(111) at $T = 4$ K. The sample was subsequently annealed at 200°C to induce the homolytic cleavage of C–Br bonds followed by radical step-growth polymerization to give *poly-1*. The polymer intermediate (Fig. 1e) exhibits a lattice periodicity that is twice the size of the molecular precursor **1**. This is consistent with the expectation that the lattice constant of the 7/9-AGNR supercell comprises a head-to-head molecular dimer. STM topography reveals a characteristic height profile and morphology that alternates between proto-9-AGNR and proto-7-AGNR segments (Fig. 1e). Annealing the sample at 300°C induces cyclodehydrogenation and leads to the fully fused 7/9-AGNR superlattices depicted in Fig. 1f, g. Bond-resolved STM images of 7/9-AGNRs were acquired by recording the out-of-phase component of constant-current dI/dV maps at low tip–sample bias using an STM tip that was spontaneously functionalized by a small molecule from the surface (Fig. 1g). Similar results were obtained for intentionally CO-functionalized tips (see Extended Data Fig. 5). A representative bond-resolved STM image depicted in Fig. 1g confirms the alternating sequence of short segments of 7-AGNRs and 9-AGNRs and the atomically precise 7/9 heterojunction interface that is characteristic of a 7/9-AGNR topological superlattice.

The local electronic structure of 7/9-AGNR superlattices was characterized using dI/dV point spectroscopy as shown in Fig. 2a. All spectra were collected after calibrating the STM tip via the well known Au(111) Shockley surface state. Spectra collected in the bulk of the 7/9-AGNR superlattice (at least 2.6 nm from a GNR end termination, corresponding to the length of one dimer unit) show a series of reproducible electronic states on both 7-AGNR and 9-AGNR segments, with peaks centred at -1.14 ± 0.07 V (peak A), -0.14 ± 0.04 V (peak B), 0.60 ± 0.04 V (peak C), and 1.61 ± 0.04 V (peak D). Since peaks B and C bracket the Fermi energy, E_F , our apparent experimental bandgap for the 7/9-AGNR superlattice is 0.74 ± 0.06 eV. This gap is substantially smaller than the experimental bandgaps measured by scanning tunnelling spectroscopy (STS) under similar conditions for both uniform 7-AGNRs (2.3 eV bandgap)²⁷ and 9-AGNRs (1.4 eV bandgap)²³ on Au(111). dI/dV maps recorded at biases corresponding to the four peak energies in Fig. 2a reveal characteristic, reproducible patterns in the local density of states (LDOS) maps for each of these four bands (Fig. 2b).

As expected for a one-dimensional topologically nontrivial system under vacuum^{28,29}, spectroscopy performed near a terminal end of the 7/9-AGNR superlattice shows markedly different behaviour compared to the bulk spectroscopy shown in Fig. 2a, b. Figure 3b reveals three new spectral features confined to the last supercell of a 7/9-AGNR superlattice that are absent in the bulk (the new states are marked end states 1–3). The dI/dV maps depicted in Fig. 3c show the characteristic LDOS patterns of end states 1–3 (in contrast, the dI/dV maps of bulk features B and C show that they are absent from the last supercell). It is notable that end state 2 lies virtually mid-gap between the bulk peaks B and C, while end states 1 and 3 lie within the A/B and C/D energy gaps, respectively. While the zigzag end termination shown in Fig. 3a is the most common 7/9-AGNR superlattice termination, the alternative termination (an armchair edge emerging from the 9-AGNR segment) was also observed and exhibits notable end-state behaviour as well (Extended Data Fig. 4).

The observed existence of end states 1–3 as well as the bulk behaviour of the 7/9-AGNR superlattice follow the predictions of ref.¹¹ since each of the two new interface-state-derived bands (B and C) can be shown to have a Zak phase equal to zero for the terminal geometry in Fig. 3a, making the system topologically nontrivial for all three gaps (A/B, B/C and C/D). To quantitatively verify the topological origins of the local electronic structure, we first compare the measurements to simulations performed using first-principles density functional theory (DFT) within the local-density approximation (LDA). Figure 2d shows the theoretical bulk density of states (DOS) for a freestanding 7/9-AGNR superlattice. A series of peaks arise from the superlattice band structure (Fig. 4c); these peaks are labelled the valence band (VB), the occupied topologically induced band (OTB), the unoccupied topologically induced band (UTB) and the conduction band (CB). The OTB and UTB are so named because they arise from the topologically induced interface states located at each internal 7/9-AGNR heterojunction. The OTB and UTB features both have a double-peak structure in the DOS plot owing to the presence of two Van Hove singularities in each band. The relative positions of these four bands correlate with peaks A–D, observed experimentally in the bulk region of a 7/9-AGNR superlattice as shown in Fig. 2a. Notably, the anomalously small bandgap observed experimentally is nicely reproduced by the DFT calculations, which predict a gap of 0.52 eV (see band structure in Fig. 4c). It is not surprising that this value is smaller than the gap observed experimentally (0.74 eV) given that DFT tends to underestimate quasiparticle bandgaps^{18,30}, even

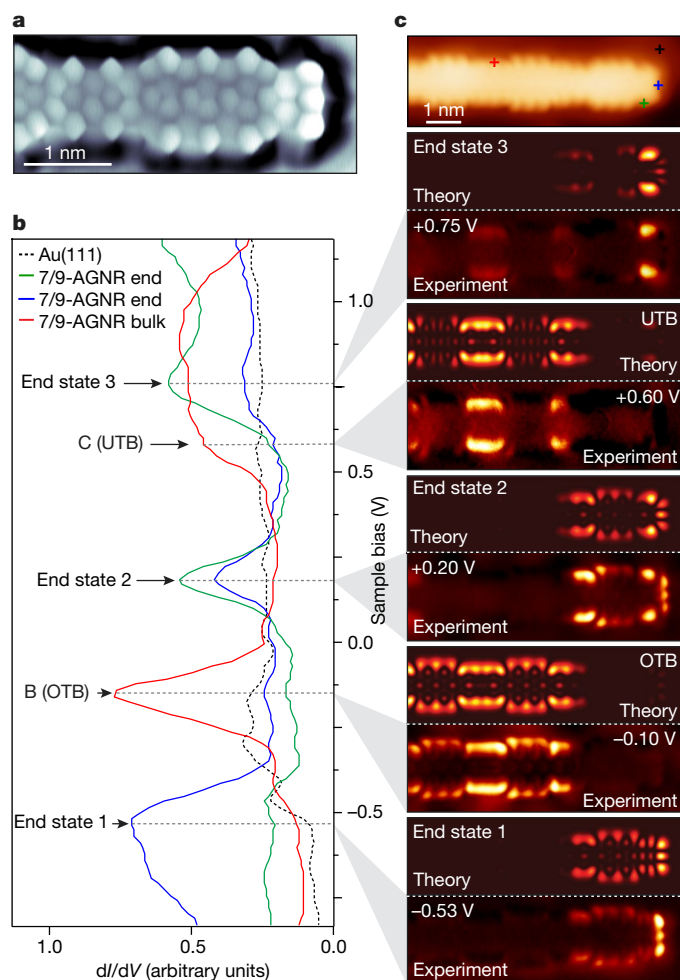


Fig. 3 | Electronic structure of 7/9-AGNR superlattice end states.

a, Bond-resolved STM image of the most common 7/9-AGNR superlattice end termination ($V_s = 0.02$ V, $I_t = 80$ pA, $f = 581$ Hz, $V_{ac} = 12$ mV). **b**, The red (green, blue) curve shows dI/dV point spectroscopy collected above the 7/9-AGNR superlattice bulk (end). The dotted black curve shows the spectrum collected on bare Au(111). Spectroscopy parameters: $V_{ac} = 20$ mV, $f = 581$ Hz. **c**, STM topography showing the tip position for STS in **b** (marked by a plus sign) ($V_s = -0.10$ V, $I_t = 80$ pA). Experimental dI/dV map (bottom) compared to the corresponding theoretical LDOS maps (top) for end state 3, UTB, end state 2, OTB, and end state 1. The dI/dV map parameters are $I_t = 50$ pA, $V_{ac} = 20$ mV and $f = 581$ Hz. All STM data were obtained at $T = 4$ K. Corresponding DFT-calculated LDOS maps are simulated at a height of 4 Å above a freestanding 7/9-AGNR superlattice comprised of eight supercells (to view the theoretical energy-dependent DOS plot, see Extended Data Fig. 3).

accounting for the screening effects of the underlying Au substrate²⁶. Figure 2c shows that the theoretical LDOS maps at 4 Å above the plane of the 7/9-AGNR superlattice at energies corresponding to the VB, the OTB, the UTB and the CB. These LDOS maps are in excellent agreement with the experimental LDOS patterns shown in Fig. 2b. This agreement between *ab initio* theory and experiment confirms that peaks A–D observed in STS do indeed originate from the intrinsic VB, OTB, UTB and CB of the GNR superlattice.

The topological origin of the 7/9-AGNR superlattice bulk electronic properties is further indicated by fitting equation (1) to the UTB and OTB band structure of Fig. 4c, which yields $t_1 = 0.33$ eV (for hopping across 9-AGNR segments) and $t_2 = -0.07$ eV (for hopping across 7-AGNR segments). The hopping terms have opposite signs, which is consistent with a direct gap at the Γ point. The stronger hopping term across the 9-AGNR segment arises from its smaller intrinsic bandgap, which allows the interface state to extend further into it and to overlap more strongly with adjacent interface states (Fig. 4a). This overlap

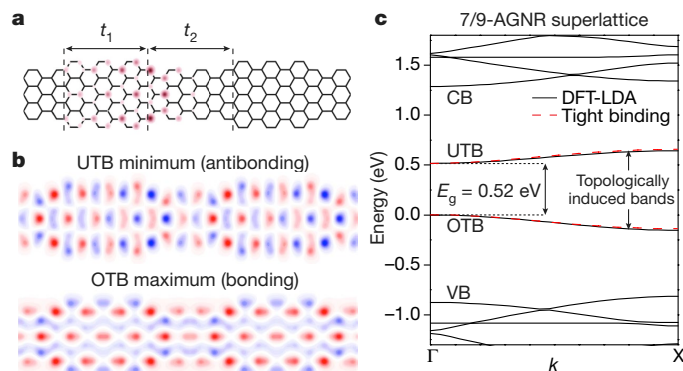


Fig. 4 | Topological bands in 7/9-AGNR superlattice. **a**, Wireframe sketch of the 7/9-AGNR superlattice with superimposed charge density for only a single, isolated interface state. Arrows represent hopping amplitudes that couple interface states through different AGNR segments. **b**, DFT-LDA wavefunctions for the OTB maximum and UTB minimum are composed of bonding and antibonding linear combinations of adjacent interface-state wavefunctions. The wavefunctions are plotted in the plane 1 Å above the GNR atomic plane to demonstrate bonding symmetry. **c**, Solid black curves show the DFT-LDA band structure of a freestanding 7/9-AGNR superlattice. The dashed red curve shows a tight-binding fit to DFT OTB and UTB using equation (1).

causes the Bloch wavefunctions of the OTB and UTB at the Γ point to reflect, respectively, bonding and anti-bonding interface states coupled through 9-AGNR segments (Fig. 4b). The presence of these topological interface-state-derived bands contrasts with the band structure of a nearly structurally equivalent, but topologically trivial, 7/9-AGNR superlattice (Extended Data Fig. 2), which completely lacks the two interface-state-derived bands owing to the absence of variation in the value of \mathbb{Z}_2 along its length. The substantial bandgap reduction seen in our 7/9-AGNR superlattice compared to the properties of individual 7-AGNRs and 9-AGNRs thus arises from the controlled incorporation of topological interface states into this bottom-up system.

The end-state properties of the 7/9-AGNR superlattice can be understood by examining the overall \mathbb{Z}_2 value of the system for successive band occupation up to a particular bandgap. For the supercell associated with the experimentally observed end structure shown in Fig. 3a, the occupation of bands up to and including the VB results in the system being topologically nontrivial (that is, $\mathbb{Z}_2 = 1$; Extended Data Fig. 3), and thus requires the existence of a 7/9-AGNR/vacuum interface state in the VB/OTB energy gap (that is, the experimental state labelled ‘end state 1’ in Fig. 3b). The behaviour in the next OTB/UTB energy gap is determined by the Zak phase of the OTB plus those of the entire band complex below it. Although the OTB and UTB arise directly from coupled topological interface states, analysis of the Zak phase of these bands shows that it is zero (topologically trivial) for each band for the terminating geometry considered (Extended Data Fig. 3). The overall value of \mathbb{Z}_2 thus remains $\mathbb{Z}_2 = 1$ for the OTB/UTB and UTB/CB bandgaps, making the existence of topological 7/9-AGNR/vacuum end states required in both energy gaps, just as seen experimentally (that is, end states 2 and 3 in Fig. 3b). Similar analysis reveals nontrivial topology for the other, less common, experimentally observed superlattice end structure (Extended Data Fig. 4).

This topological behaviour can also be clearly seen in our simulations of the end region of a 7/9-AGNR superlattice calculated for a finite 7/9-AGNR consisting of eight supercells. The LDOS of this structure exactly reproduces end states 1–3 in the three energy gaps, as discussed above (Extended Data Fig. 3). A direct comparison between the experimental dI/dV maps and the calculated LDOS maps of end states 1–3 shows excellent agreement between theory and experiment (Fig. 3c). Similarly, the experimental dI/dV maps and the calculated LDOS maps of the OTB and UTB show high intensity throughout the bulk 7/9-AGNR superlattice but decay rapidly in the last supercell that terminates the GNR.

In conclusion, we have demonstrated that it is possible to rationally engineer both the local and global GNR electronic topology via careful design of molecular precursors used in bottom-up synthesis. This approach enables the deterministic design of topological interface states both in the GNR bulk as well as in the GNR/vacuum termination region. Superlattices of topological interface states allow the formation of new bulk frontier bands (the OTB and UTB) that are energetically distinct from the intrinsic band structures associated with the parent 7- and 9-AGNRs. In principle, the properties of these topologically induced bands can be fine-tuned through topology-conserving modification of the superlattice components to create effective antiferromagnetic Heisenberg spin-1/2 chains with robust spin centres at each internal 7/9-AGNR interface¹¹. If placed in close proximity to a superconductor, the ends of these antiferromagnetic chains are predicted to host Majorana fermion states³¹.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0376-8>.

Received: 1 February 2018; Accepted: 7 June 2018;

Published online 8 August 2018.

- Kane, C. L. & Mele, E. J. Quantum spin hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005).
- Kane, C. L. & Mele, E. J. Z_2 topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95**, 146802 (2005).
- Fu, L. Topological crystalline insulators. *Phys. Rev. Lett.* **106**, 106802 (2011).
- Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
- Wu, C., Bernevig, B. A. & Zhang, S.-C. Helical liquid and the edge of quantum spin Hall systems. *Phys. Rev. Lett.* **96**, 106401 (2006).
- Xu, C. & Moore, J. E. Stability of the quantum spin Hall effect: effects of interactions, disorder, and Z_2 topology. *Phys. Rev. B* **73**, 045322 (2006).
- Hsieh, D. et al. A topological Dirac insulator in a quantum spin Hall phase. *Nature* **452**, 970 (2008).
- Hsieh, D. et al. Observation of unconventional quantum spin textures in topological insulators. *Science* **323**, 919–922 (2009).
- Teo, J. C. Y., Fu, L. & Kane, C. L. Surface states and topological invariants in three-dimensional topological insulators: application to $\text{Bi}_{1-x}\text{Sb}_x$. *Phys. Rev. B* **78**, 045426 (2008).
- Nishide, A. et al. Direct mapping of the spin-filtered surface bands of a three-dimensional quantum spin Hall insulator. *Phys. Rev. B* **81**, 041309 (2010).
- Cao, T., Zhao, F. & Louie, S. G. Topological phases in graphene nanoribbons: junction states, spin centers, and quantum spin chains. *Phys. Rev. Lett.* **119**, 076401 (2017).
- Zak, J. Berry's phase for energy bands in solids. *Phys. Rev. Lett.* **62**, 2747 (1989).
- Cheon, S., Kim, T.-H., Lee, S.-H. & Yeom, H. W. Chiral solitons in a coupled double Peierls chain. *Science* **350**, 182 (2015).
- Kim, T.-H., Cheon, S. & Yeom, H. W. Switching chiral solitons for algebraic operation of topological quaternary digits. *Nat. Phys.* **13**, 444–447 (2017).
- Brazovskii, S., Brun, C., Wang, Z.-Z. & Monceau, P. Scanning-tunneling microscope imaging of single-electron solitons in a material with incommensurate charge-density waves. *Phys. Rev. Lett.* **108**, 096801 (2012).
- Su, W. P., Schrieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
- Son, Y.-W., Cohen, M. L. & Louie, S. G. Energy gaps in graphene nanoribbons. *Phys. Rev. Lett.* **97**, 216803 (2006).
- Yang, L., Park, C.-H., Son, Y.-W., Cohen, M. L. & Louie, S. G. Quasiparticle energies and band gaps in graphene nanoribbons. *Phys. Rev. Lett.* **99**, 186801 (2007).

- Cai, J. et al. Atomically precise bottom-up fabrication of graphene nanoribbons. *Nature* **466**, 470 (2010).
- Nguyen, G. D. et al. Atomically precise graphene nanoribbon heterojunctions from a single molecular precursor. *Nat. Nanotechnol.* **12**, 1077 (2017).
- Chen, Y.-C. et al. Molecular bandgap engineering of bottom-up synthesized graphene nanoribbon heterojunctions. *Nat. Nanotechnol.* **10**, 156–160 (2015).
- Bronner, C. et al. Aligning the band gap of graphene nanoribbons by monomer doping. *Angew. Chem. Int. Ed.* **52**, 4422–4425 (2013).
- Talirz, L. et al. On-surface synthesis and characterization of 9-atom wide armchair graphene nanoribbons. *ACS Nano* **11**, 1380–1388 (2017).
- Vo, T. H. et al. Nitrogen-doping induced self-assembly of graphene nanoribbon-based two-dimensional and three-dimensional metamaterials. *Nano Lett.* **15**, 5770–5777 (2015).
- Kimouche, A. et al. Ultra-narrow metallic armchair graphene nanoribbons. *Nat. Commun.* **6**, 10177 (2015).
- Cai, J. et al. Graphene nanoribbon heterojunctions. *Nat. Nanotechnol.* **9**, 896 (2014).
- Ruffieux, P. et al. Electronic structure of atomically precise graphene nanoribbons. *ACS Nano* **6**, 6930–6935 (2012).
- Rhim, J.-W. et al. Bulk-boundary correspondence from the intercellular Zak phase. *Phys. Rev. B* **95**, 035421 (2017).
- Grusdt, F. et al. Topological edge states in the one-dimensional superlattice Bose-Hubbard model. *Phys. Rev. Lett.* **110**, 260405 (2013).
- Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B* **34**, 5390–5413 (1986).
- Heimes, A., Kotetes, P. & Schön, G. Majorana fermions from Shiba states in an antiferromagnetic chain on top of a superconductor. *Phys. Rev. B* **90**, 060507 (2014).

Acknowledgements Research supported by the Office of Naval Research MURI Program N00014-16-1-2921 (precursor design, STM spectroscopy, band structure), by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES) under award number DE-SC0010409 (precursor synthesis and characterization) and the Nanomachine Program award number DE-AC02-05CH11231 (surface growth, heterojunction analysis), by the Center for Energy Efficient Electronics Science NSF Award 0939514 (end state modelling), and by the National Science Foundation under grant DMR-1508412 (development of theory formalism). Computational resources have been provided by the DOE at Lawrence Berkeley National Laboratory's NERSC facility and by the NSF through XSEDE resources at NICS. We acknowledge useful discussions with O. Gröning from EMPA Federal Laboratories for Materials Testing and Research.

Reviewer information Nature thanks T. Heine, I. Swart and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.J.R., G.V., T. Cao, S.G.L., M.F.C. and F.R.F. initiated and conceived the research, G.V. and F.R.F. designed, synthesized and characterized the molecular precursors, D.J.R., C.B., H.R., T. Chen and M.F.C. performed on-surface synthesis and STM characterization and analysis, T.Cao., F.Z. and S.G.L. performed the DFT calculations and the theoretical analysis that predicted and interpreted the STM data. D.J.R., T. Cao, G.V., S.G.L., M.F.C. and F.R.F. wrote the manuscript. All authors contributed to the scientific discussion.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0376-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0376-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.G.L. or M.F.C. or F.R.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Precursor synthesis and GNR superlattice growth. Full details of the synthesis and characterization of **1–4** are given in the Supplementary Information. 7/9-AGNR superlattices were grown on a Au(111) single crystal under ultrahigh-vacuum conditions. Atomically clean Au(111) surfaces were prepared through iterative argon ion (Ar^+) sputter/anneal cycles. Sub-monolayer coverage of **1** on atomically clean Au(111) was obtained by sublimation using a Knudsen cell evaporator that was built in our laboratory for 20–30 min at crucible temperatures of 200–215 °C. After deposition, the surface temperature was ramped slowly ($\leq 2 \text{ K min}^{-1}$) to 200 °C and held at this temperature for 30 min to induce the radical-step growth polymerization, then ramped slowly ($\leq 2 \text{ K min}^{-1}$) to 300 °C and held there for 30 min to induce cyclodehydrogenation.

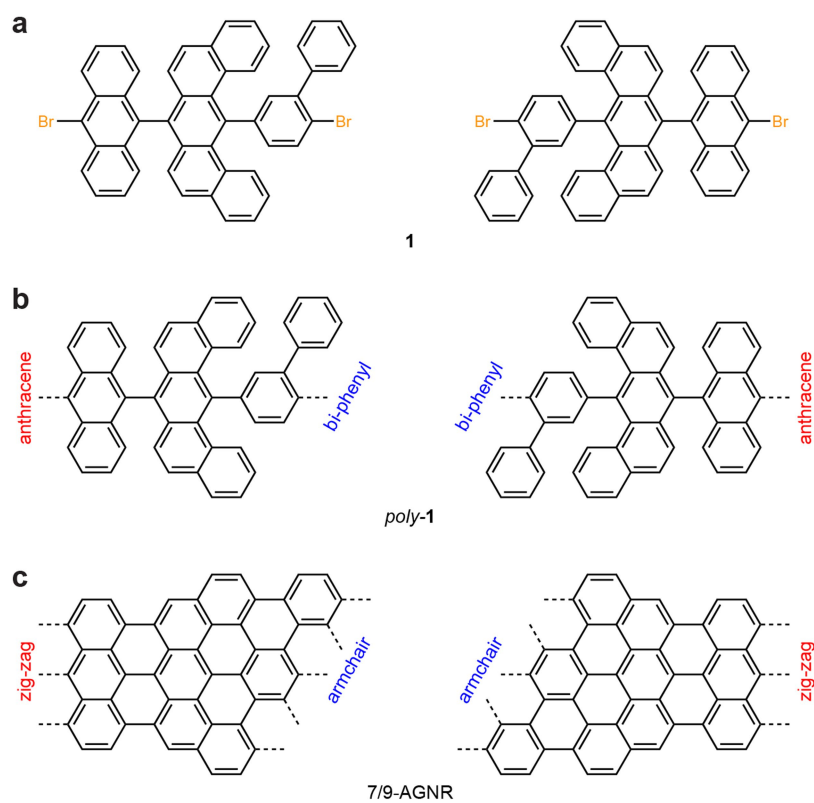
STM measurements. All STM experiments were performed using a commercial Createc LT-STM held at $T \approx 4 \text{ K}$ using platinum–iridium STM tips. All scanning probe images were edited using WSxM software³². dI/dV measurements were recorded using a lock-in amplifier with a modulation frequency of 581 Hz and a bias modulation amplitude of $V_{ac} = 10\text{--}20 \text{ mV}$. dI/dV point spectra were recorded under open feedback loop conditions. dI/dV maps were collected under constant current conditions. Bond-resolved STM images were obtained by mapping the out-of-phase dI/dV signal collected during a low bias (20 mV) dI/dV map. Peak positions in dI/dV point spectroscopy were determined by fitting the spectra with Lorentzian peaks. Each peak position is based on an average of approximately 80 spectra collected on 15 GNRs with 17 different tips, all of which were first

calibrated to the Au(111) Shockley surface state. All tips calibrated in this manner reproduce the characteristic LDOS patterns of each state in dI/dV maps (both bulk states and end states). The bulk state STS was independent of the precise AGNR terminating geometry. Similarly, the end-state STS did not depend on whether both terminal geometries were the same.

Calculations. First-principles calculations of GNR superlattices were performed using DFT in the LDA as implemented in the Quantum Espresso package^{33,34}. A supercell arrangement was used with vacuum regions carefully tested to avoid interactions between the superlattice and its periodic image. We used norm-conserving pseudopotentials with a planewave energy cut-off of 60 Ry. The structure was fully relaxed until the force on each atom was smaller than 0.02 eV \AA^{-1} . All σ dangling bonds on the edges and the ends of the GNR were capped by hydrogen atoms. A Gaussian broadening of 0.05 eV was used in the evaluation of DOS.

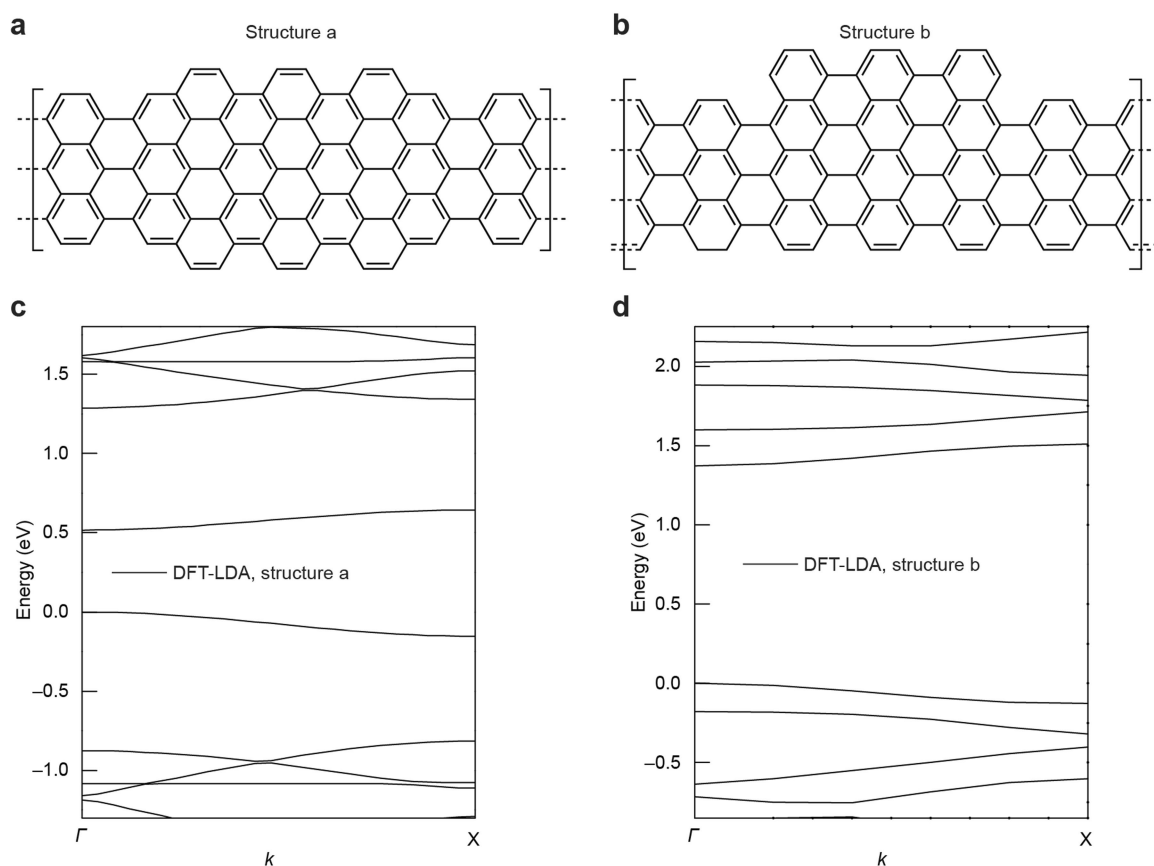
Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

32. Horcas, I. et al. WSxM: a software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.* **78**, 013705 (2007).
33. Giannozzi, P. et al. Advanced capabilities for materials modelling with QUANTUM ESPRESSO. *J. Phys. Condens. Matter* **29**, 465901 (2017).
34. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).



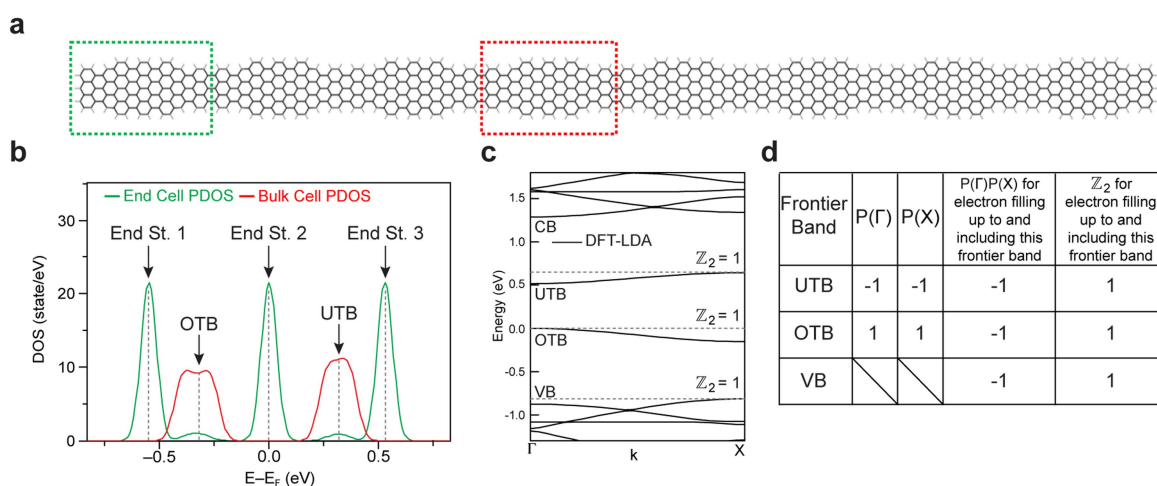
Extended Data Fig. 1 | Sterically enforced site-selective polymerization.
a, Molecular precursor **1**. **b**, Sterically distinct reaction sites during radical chain growth polymerization of *poly-1* (anthracene versus biphenyl).

c, The corresponding edge structures in the fully formed GNR (armchair and zigzag termination, respectively).



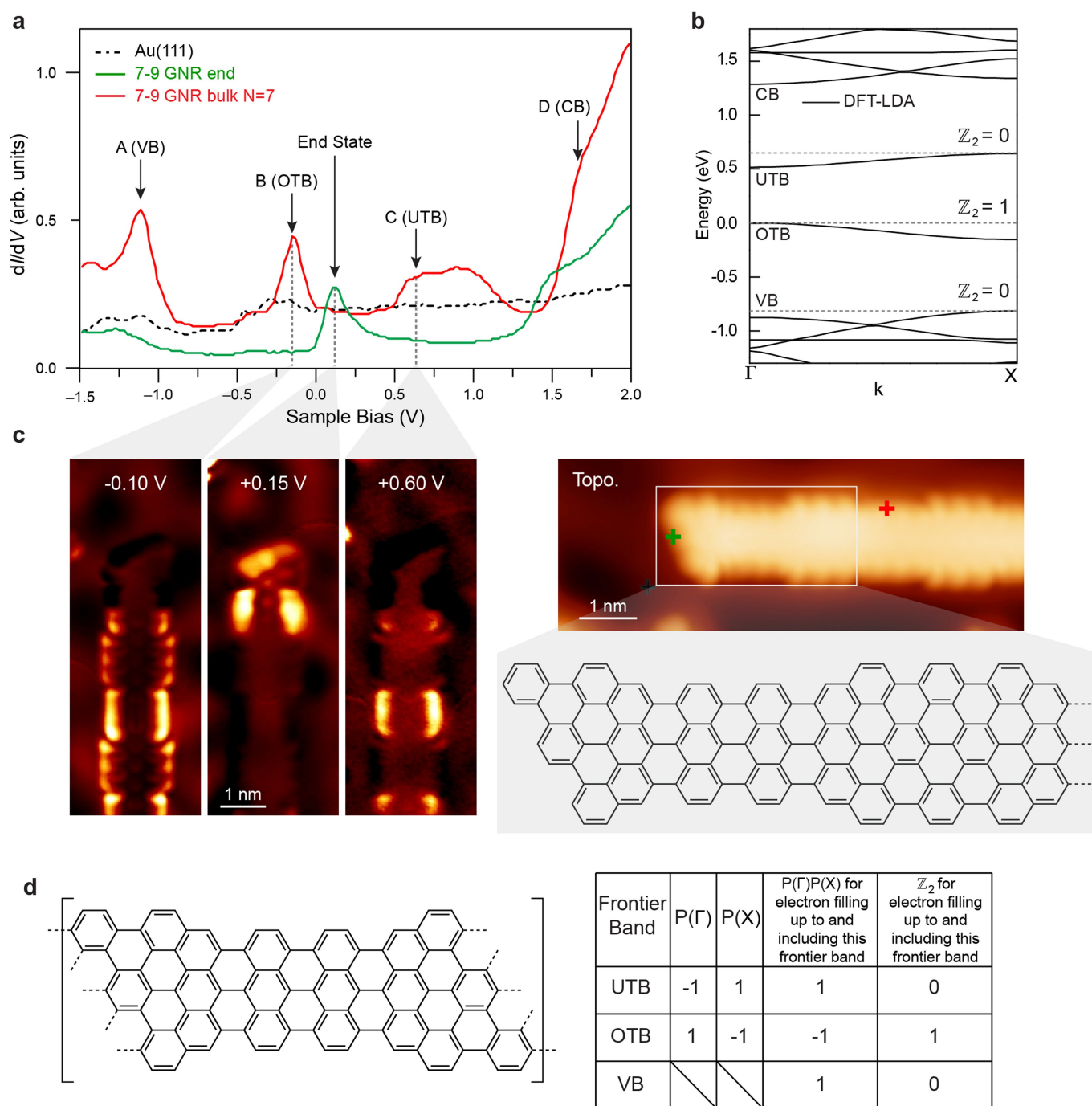
Extended Data Fig. 2 | Electronic structure of two different 7/9-AGNR superlattices. a, c, The unit cell of structure a (a) and the DFT-calculated band structure for the 7/9-AGNR superlattice composed of topologically nontrivial interfaces (c) show two new topologically induced bands.

b, d, The unit cell of structure b (b) and the DFT-calculated band structure of the 7/9-AGNR superlattice composed of topologically trivial interfaces (d) show no topologically induced bands in the energy gap region.



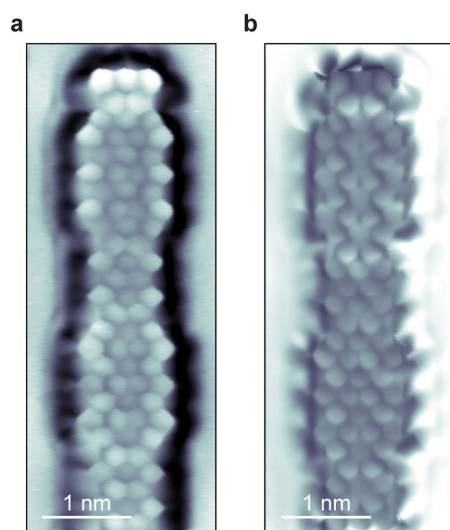
Extended Data Fig. 3 | Electronic structure of finite nontrivial 7/9-AGNR superlattice. **a**, Fully relaxed finite (8-supercell) 7/9-AGNR superlattice with end (green) and bulk (red) unit cells indicated. **b**, DFT-calculated projected DOS of the finite (8-supercell) 7/9-AGNR superlattice obtained from the end unit cell (green) and a bulk unit cell (red) (Gaussian broadening of 0.05 eV was used here). Three end states are seen that closely correspond to the experimental end states shown in Fig. 3b. **c**, DFT-calculated band structure of 7/9-AGNR showing the overall value

of \mathbb{Z}_2 for occupation up to all three energy gaps around E_F based on the edge structure shown in Fig. 3a. **d**, Chart of frontier band parity eigenvalues and corresponding \mathbb{Z}_2 invariants for electron filling up to and including a given frontier band. This superlattice end-state behaviour is different from the behaviour of a 'straight-edge' topologically nontrivial AGNR owing to the presence of multiple energy gaps that can accommodate topologically protected end states rather than only a single gap.



Extended Data Fig. 4 | Topologically nontrivial 7/9-AGNR with different end structure. **a**, The red (green) curve shows dI/dV point spectroscopy data collected on the 7/9-AGNR superlattice bulk (end) region. The dashed black curve shows the spectrum collected on bare Au(111). Only one end state is observed that lies in the energy gap between the OTB and UTB. The spectroscopy parameters are $V_{ac} = 20$ mV and $f = 581$ Hz. **b**, The DFT-calculated band structure of 7/9-AGNR shows an overall topologically nontrivial phase for the edge structure shown in **c** only for bands filled up to and including the OTB. **c**, Sketch of GNR structure and STM topographic image of additional 7/9-AGNR end terminus that is seen for <10% of all 7/9 AGNR superlattices in the

experiment. Experimental topography and dI/dV maps are shown for the OTB, the end state and the UTB (topography $V_s = -0.10$ V, $I_t = 50$ pA; dI/dV maps $I_t = 50$ pA, $f = 581$ Hz, $V_{ac} = 20$ mV). **d**, Unit cell commensurate with uncommon end terminus shown in **c**, along with chart of frontier band parity eigenvalues and corresponding \mathbb{Z}_2 invariants for electron filling up to and including a given frontier band. For this edge structure only the OTB/UTB gap is topologically nontrivial and supports a topologically protected end state. The UTB/CB and VB/OTB gaps are topologically trivial and do not support end states, unlike the behaviour seen for the other, more common, termination shown in Fig. 3a.



Extended Data Fig. 5 | Comparison of bond-resolved STM measurements conducted with CO-functionalized and spontaneously functionalized STM tips. a, Bond-resolved STM image of 7/9-AGNR, obtained with a tip that was spontaneously functionalized via an unknown molecule from the surface ($V_s = 0.02$ V, $I_t = 80$ pA, $f = 581$ Hz, $V_{ac} = 12$ mV). **b,** Bond-resolved STM image of 7/9-AGNR, obtained with a tip deliberately functionalized with CO ($V_s = 0.02$ V, $I_t = 180$ pA, $f = 581$ Hz, $V_{ac} = 12$ mV).

Engineering of robust topological quantum phases in graphene nanoribbons

Oliver Gröning^{1,7*}, Shiyong Wang^{1,2,7}, Xuelin Yao^{3,7}, Carlo A. Pignedoli¹, Gabriela Borin Barin¹, Colin Daniels⁴, Andrew Cupo⁴, Vincent Meunier⁴, Xinliang Feng⁵, Akimitsu Narita³, Klaus Müllen³, Pascal Ruffieux¹ & Roman Fasel^{1,6}

Boundaries between distinct topological phases of matter support robust, yet exotic quantum states such as spin–momentum locked transport channels or Majorana fermions^{1–3}. The idea of using such states in spintronic devices or as qubits in quantum information technology is a strong driver of current research in condensed matter physics^{4–6}. The topological properties of quantum states have helped to explain the conductivity of doped *trans*-polyacetylene in terms of dispersionless soliton states^{7–9}. In their seminal paper, Su, Schrieffer and Heeger (SSH) described these exotic quantum states using a one-dimensional tight-binding model^{10,11}. Because the SSH model describes chiral topological insulators, charge fractionalization and spin–charge separation in one dimension, numerous efforts have been made to realize the SSH Hamiltonian in cold-atom, photonic and acoustic experimental configurations^{12–14}. It is, however, desirable to rationally engineer topological electronic phases into stable and processable materials to exploit the corresponding quantum states. Here we present a flexible strategy based on atomically precise graphene nanoribbons to design robust nanomaterials exhibiting the valence electronic structures described by the SSH Hamiltonian^{15–17}. We demonstrate the controlled periodic coupling of topological boundary states¹⁸ at junctions of graphene nanoribbons with armchair edges to create quasi-one-dimensional trivial and non-trivial electronic quantum phases. This strategy has the potential to tune the bandwidth of the topological electronic bands close to the energy scale of proximity-induced spin–orbit coupling¹⁹ or superconductivity²⁰, and may allow the realization of Kitaev-like Hamiltonians³ and Majorana-type end states²¹.

The fundamental features of the SSH model—which describes a one-dimensional chain of dimerized, coupled and spinless fermion states—are summarized in Fig. 1. Conceptually, its basic elements are an ensemble of equivalent fermion states $|\psi_i\rangle$ at each site i of the chain, an intra-cell coupling t_n between two such states within the same dimer, and an inter-cell coupling t_m between states of neighbouring dimers (Fig. 1a). The corresponding spinor-based Hamiltonian $H(k) = d_x(k)\sigma_x + d_y(k)\sigma_y$ (with $d_x(k) = t_n + t_m \cos(k)$ and $d_y(k) = t_m \sin(k)$) leads to the energy spectrum¹¹ $E(k) = \pm \sqrt{t_n^2 + t_m^2 + 2t_n t_m \cos(k)}$. This dispersion relation yields three extremal phases: (i) an intra-cell decoupled, insulating phase with $E(k) = \pm t_m$ for $t_n = 0$ and $t_m \neq 0$; (ii) a metallic phase with $E(\pi) = 0$ and $E(0) = \pm 2t_n$ for equal coupling strengths $t_n = t_m \neq 0$; and (iii) an inter-cell decoupled, insulating phase with $E(k) = \pm t_n$ for $t_n \neq 0$ and $t_m = 0$.

These three extremal solutions of the SSH chain can be smoothly connected by introducing a phase factor $\phi \in [0, \pi/2]$ governing the strength of t_n and t_m via $t_n = \gamma \sin^2(\phi)$ and $t_m = \gamma \cos^2(\phi)$, where γ denotes the bandwidth. The corresponding series of band structures $E(k, \phi)$ in Fig. 1b reveals non-dispersive band structures (orange) for the two insulating chain configurations at $\phi = 0$ and $\phi = \pi/2$, while for

$\phi = \pi/4$ (blue) a gapless metallic phase is found. That the smooth transition between two insulating phases can only occur by closing the gap is clear evidence of their distinct topological class. This class can be assigned using the winding number of $\mathbf{r}(k, \phi) = (d_x(k, \phi), d_y(k, \phi))$ around the origin as a \mathbb{Z}_2 topological invariant¹¹, which is $\mathbb{Z}_2 = 1$ for $\phi < \pi/4$ and $t_n < t_m$, making the corresponding phases topologically non-trivial, and topologically trivial with $\mathbb{Z}_2 = 0$ for $\phi > \pi/4$ and $t_n > t_m$. Unfortunately, the winding number cannot be directly determined in experiments. However, the bulk–boundary correspondence, that is, the relation between the bulk winding number and the existence or absence of boundary states, offers a convenient experimental approach with which to determine a topological class. In the energy spectrum of a finite SSH chain of 25 dimers (Fig. 1c) the topologically non-trivial phases for $\phi < \pi/4$ can readily be distinguished from the trivial ones with $\phi > \pi/4$ by the presence of two degenerate zero-energy states localized at the chain ends.

Specifically designed graphene nanoribbons (GNRs) provide a platform with which to realize a class of robust solid-state nanomaterials that can flexibly encompass all three of the abovementioned quantum phases of the SSH chain. The atomically precise structural control required to rationally engineer the corresponding electronic structures can be achieved by on-surface synthesis²². Since the first successful bottom-up synthesis of GNRs by polymerization of dedicated molecular precursors¹⁵, a wide variety of GNRs exhibiting different width, chirality, edge structure and chemical doping has been realized^{16,17}. The chemical robustness of GNRs allows their handling under ambient conditions²³ and their integration into high-performance electronic nanodevices²⁴, promising a technological exploitation of GNR-based topological quantum phases¹⁸.

The ability to flexibly engineer SSH-like topological quantum phases in GNRs requires a suitable electronic state representing $|\psi_i\rangle$. We identify such a state in the zero-energy boundary state at the junction of two armchair graphene nanoribbons (N -AGNR) of different widths. Here N denotes the number of transverse carbon atom rows¹⁵. The boundary state we are considering here is itself of topological origin¹⁸. To understand this, we consider that N -AGNRs can be classified into three families according to their electronic properties. For $N = 3p$ and $N = 3p + 1$ (where p is an integer) the corresponding AGNRs exhibit a gapped electronic structure, whereas for $N = 3p + 2$ a gapless (that is, metallic) behaviour is observed at the tight-binding level of theory²⁵. At a smooth junction between a gapped N -AGNR with $N = 3p + 1$ and a gapped $N = 3p + 3$ AGNR (that is, with two additional rows of carbon atoms) (see Supplementary Figs. 1–4), a zero-energy boundary state occurs owing to the gapless $N = 3p + 2$ intermediate (Supplementary Figs. 5–9). This situation is analogous to polyacetylene, where the smooth transition from one bond alternation pattern to the complementary one can only proceed via closure of the gap, leading to a zero-energy soliton state^{8,11}. The wavefunction of the corresponding boundary state at a 7-AGNR/9-AGNR junction is displayed in Fig. 1d.

¹Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland. ²School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China. ³Max Planck Institute for Polymer Research, Mainz, Germany. ⁴Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY, USA. ⁵Department of Chemistry and Food Chemistry, Technische Universität Dresden, Dresden, Germany. ⁶Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland. ⁷These authors contributed equally: Oliver Gröning, Shiyong Wang, Xuelin Yao. *e-mail: oliver.groening@empa.ch

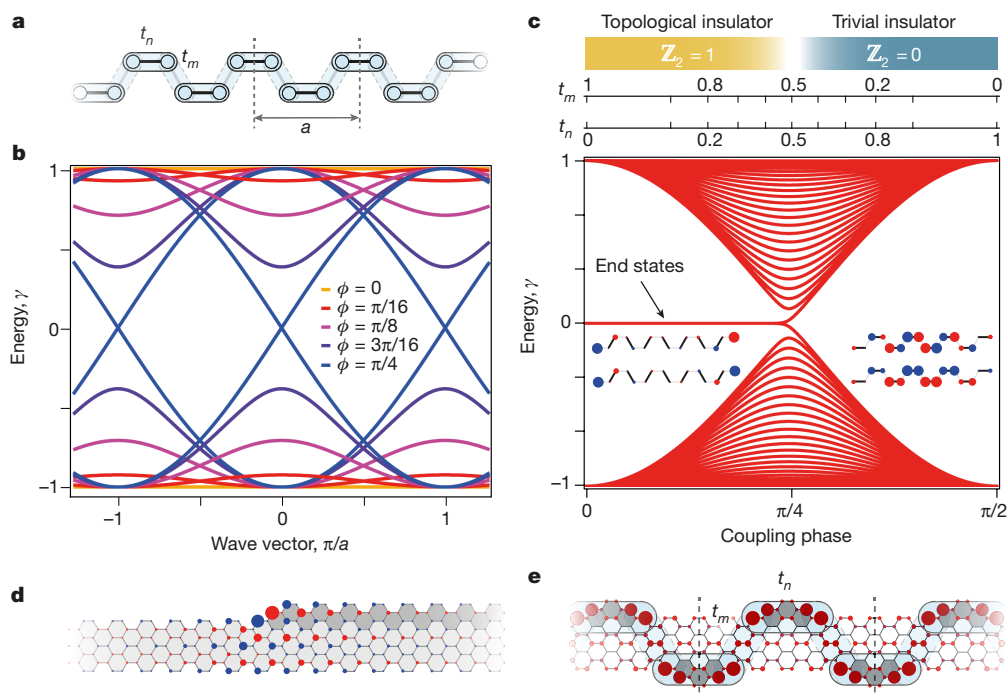


Fig. 1 | The SSH model and its realization in edge-extended graphene nanoribbons. **a**, Schematic representation of the dimerized *cis*-polyacetylene-like SSH chain with illustration of the intra-cell coupling t_n and inter-cell coupling t_m within and between dimers, respectively (where a denotes the unit cell size). **b**, Dispersion relation $E(k)$ of the SSH chain displayed in **a** as a function of the phase factor ϕ , which governs the coupling strengths t_n and t_m (see text). **c**, Energy level diagram as a function of ϕ for a finite SSH chain of 25 dimers, revealing topological zero-energy modes for $\phi < \pi/4$ (that is, $t_n < t_m$). Also shown is the wavefunction of the frontier orbitals (that is, closest to $E = 0$ eV) for a

short 8-dimer chain with localized end-state character for $\phi < \pi/4$ and extended bulk-like character for $\phi > \pi/4$. **d**, Wavefunction of the N -AGNR to $(N + 2)$ -AGNR boundary state at an isolated smooth junction between 7-AGNR and 9-AGNR. The size of the circles in **d** denotes wavefunction amplitude and colour (blue to red) indicates wavefunction parity. **e**, Schematic representation of the frontier orbitals (size of circles denotes charge density) of a 7-AGNR with staggered edge extensions leading to short 9-AGNR segments. The corresponding 7-AGNR to 9-AGNR boundary states couple within the 9-AGNR segments (t_n) and across the 7-AGNR backbone (t_m), analogously to the *cis*-SSH chain illustrated in **a**.

Creating a periodic sequence of such boundary states along and across the N -AGNR backbone, by local extension to a finite $(N + 2)$ -AGNR segment (Fig. 1e), produces an effective solid-state analogue of a *cis*-SSH chain. Here, the index n denotes the length of the $(N + 2)$ -AGNR segment and m corresponds to the separation between the opposite segments across the backbone. The resulting staggered (S) ribbon structure is labelled N -AGNR- $S(n, m)$. Thereby, the structure shown in Fig. 1e and Fig. 2b with $N = 7$, $n = 1$ and $m = 3$ is denoted as 7-AGNR- $S(1, 3)$ (see Supplementary Figs. 1–4 for details). In terms of the SSH Hamiltonian, n is directly related to the intra-cell coupling t_n while m determines the inter-cell coupling t_m .

The tight-binding bulk band structure of the staggered 7-AGNR- $S(1, 3)$ is compared to the band structure of the pristine 7-AGNR backbone in Fig. 2. The appearance of four dispersive bands around the Fermi energy of the 7-AGNR backbone structure is readily observed (see also Supplementary Figs. 10 and 11). These bands are in excellent agreement with the zone-folded SSH energy spectrum $E(k)$ (blue solid lines in Fig. 2b) with $t_n = 0.45$ eV and $t_m = 0.59$ eV.

We present a synthetic design to experimentally realize the staggered 7-AGNR- $S(1, 3)$ structure by using 6,11-bis(10-bromoanthracen-9-yl)-1,4-dimethyltetracene (BADMT, monomer **1**) as precursor monomer. The methyl groups can form zigzag edges smoothly bridging the 7- and 9-AGNR segments via cyclization with the neighbouring aromatic rings, forming the intermediate 8-AGNR structure. The corresponding on-surface synthesis route (Fig. 2c) consists of the sublimation of monomer **1** onto a clean Au(111) surface, subsequent thermal precursor activation (dehalogenation) and polymerization at 200 °C, and finally cyclodehydrogenation of the polymer at 400 °C. A constant-height non-contact atomic force microscopy (nc-AFM) image of the resulting structure is shown in Fig. 2d. The chemical stability of this GNR was investigated by Raman spectroscopy (Supplementary

Fig. 27), and no spectral changes were detected after 5 days under ambient conditions, consistent with the high stability of the pristine backbone 7-AGNR²³.

STS investigation reveals that the 2.4-eV bandgap of the pristine 7-AGNR on Au(111)^{26,27} is drastically reduced to 0.65 ± 0.1 eV for the 7-AGNR- $S(1, 3)$. Constant-current dI/dV maps of the main spectroscopic features around the gap (Fig. 2e) can be reliably assigned to the bottom and the top of the valence band (VB) and conduction band (CB), respectively, by comparison with tight-binding simulations (Fig. 2f). The experimentally observed total bandwidth $\Delta E_{\text{exp}} = 1.6$ eV (VB minimum to CB maximum, see Supplementary Figs. 15, 17) is in good agreement with the one found from the tight-binding calculations $\Delta E_{\text{TB}} = 2\sqrt{t_n^2 + t_m^2 + 2t_n t_m} = 2.08$ eV with $t_n = 0.45$ eV and $t_m = 0.59$ eV. From density functional theory (DFT, Supplementary Fig. 12) we deduce $\Delta E_{\text{DFT}} = 1.95$ eV with $t_n = 0.37$ eV and $t_m = 0.60$ eV. The symmetry of $E(k)$ with regard to exchange of t_n and t_m does not allow us to determine which coupling term prevails and it remains an open question whether the 7-AGNR- $S(1, 3)$ structure belongs to the topologically non-trivial class ($\mathbb{Z}_2 = 1$ with $t_m > t_n$) or the topologically trivial class ($\mathbb{Z}_2 = 0$ with $t_m < t_n$).

To clarify this question we exploit the bulk–boundary correspondence¹¹ and check for the presence of end states at the termini of the N -AGNR- $S(n, m)$ nanoribbon family. There is, however, a complication arising from the concomitant presence of zigzag termini related end states of the N -AGNR backbone²⁸. Both types of end states have topological origins but of different nature. As detailed in Supplementary Figs. 18–20, these two states can interact and hybridize such that the SSH end state is no longer present at zero energy. To prevent this, the terminus of the N -AGNR- $S(n, m)$ needs to be extended by a sufficiently long segment of pristine N -AGNR backbone, as illustrated in Fig. 3a and b. The resulting local density of states (LDOS) at the end of

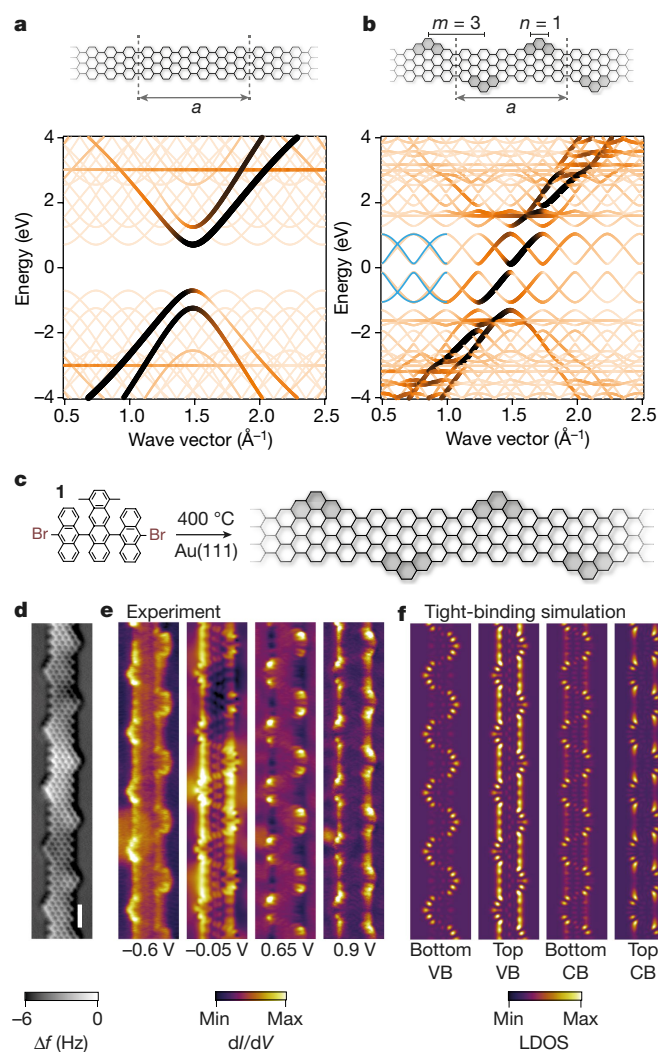


Fig. 2 | Electronic structure of the staggered edge-extended 7-AGNR-S(1,3) nanoribbon. **a, b**, Structural models and tight-binding band-structure diagrams for the pristine 7-AGNR backbone (**a**) and the staggered 7-AGNR-S(1,3) (with $\gamma_0 = 3$ eV) (**b**). The thickness of the lines and the colour—black (higher) to brown (lower)—denote the magnitude of $|\langle k | \psi_n(k) \rangle|$, that is, the projection of the electronic states onto free electron states (Brillouin zone unfolding). Solid blue lines in **b** denote the analytical bands of the *cis*-SSH chain. **c**, Schematic representation of the on-surface synthesis route from monomer **1** to 7-AGNR-S(1,3). **d**, Constant-height nc-AFM image (with CO-functionalized tip) of the frequency shift Δf of a 7-AGNR-S(1,3) segment on Au(111). **e**, STS dI/dV spectra taken at positions indicated by the markers of the corresponding colour in **d**. **f**, Experimental constant-current dI/dV maps at the top of the VB (-0.05 V, $I = 200$ pA), in the gap ($+0.25$ V, $I = 500$ pA) and at the bottom of the CB ($+0.65$ V, $I = 500$ pA) of the 7-AGNR-S(1,3) shown in **e**. **g**, Tight-binding-simulated charge-density map of the bottom of the CB, computed for the experimental structure (**d**). The 1 nm scale bar in **d** applies to all maps **d**–**f**.

the N -AGNR-S(n, m) segment (indicated by the arrows) as a function of m is shown in Fig. 3a and b for the 7-AGNR-S(1, m) and 7-AGNR-S(3, m) nanoribbon families, respectively.

The ($m = 1$) 7-AGNR-S(1,1) exhibits a zero-energy end state, indicating that it belongs to the topologically non-trivial phase ($\phi < \pi/4$) with $t_n < t_m$. Increasing m decreases t_m while t_n remains approximately constant ($n = 1$). For $m = 2$ the LDOS shows a closing of the gap corresponding to $t_n \approx t_m$ ($\phi \approx \pi/4$), thus marking the metallic intermediate separating the non-trivial 7-AGNR-S(1,1) from the trivial 7-AGNR-S(1,3), which shows a gap again but with no zero-energy end

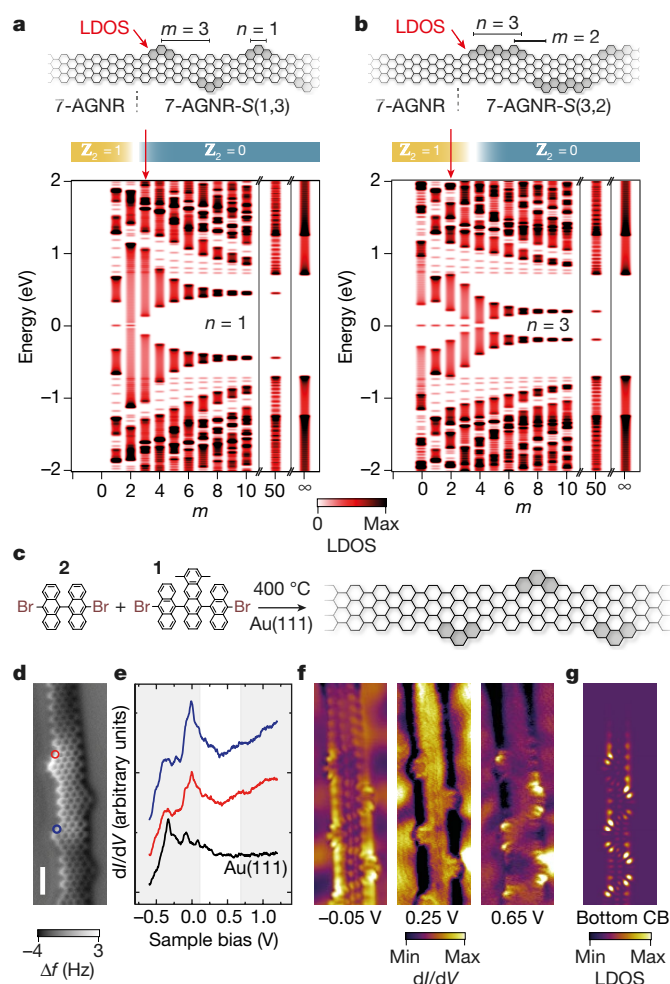


Fig. 3 | Bulk-boundary correspondence for the staggered edge-extended 7-AGNR-S(n, m) nanoribbon family. **a, b**, LDOS plots for the 7-AGNR backbone extended 7-AGNR-S(1, m) (**a**) and 7-AGNR-S(3, m) (**b**) nanoribbon families, evaluated at the end of the 7-AGNR-S(n, m) segment (see 'LDOS'-labelled red arrows). The 7-AGNR-S(1,3), whose structural model is depicted in **a**, exhibits no zero-energy end state (LDOS indicated by red arrow) and thus belongs to the topologically trivial $\mathbb{Z}_2 = 0$ class. Conversely, the 7-AGNR-S(3,2) structure (model in **b**, LDOS indicated by red arrow) reveals zero-energy end states and thus belongs to the topologically non-trivial $\mathbb{Z}_2 = 1$ class. **c**, Synthetic pathway to the 7-AGNR backbone extended 7-AGNR-S(1,3) nanoribbon using **1** and **2** as precursor molecules. **d**, Constant-height nc-AFM frequency shift (Δf) image of a 7-AGNR-S(1,3) segment (acquired with CO-functionalized tip). **e**, STS dI/dV spectra taken at positions indicated by the markers of the corresponding colour in **d**. **f**, Experimental constant-current dI/dV maps at the top of the VB (-0.05 V, $I = 200$ pA), in the gap ($+0.25$ V, $I = 500$ pA) and at the bottom of the CB ($+0.65$ V, $I = 500$ pA) of the 7-AGNR-S(1,3) shown in **e**. **g**, Tight-binding-simulated charge-density map of the bottom of the CB, computed for the experimental structure (**d**). The 1 nm scale bar in **d** applies also to panel **f** and **g**.

states. For $n = 3$ (Fig. 3b), t_n is reduced and the non-trivial to trivial transition with $t_n \approx t_m$ should occur at larger m (that is, smaller t_m) than in the $n = 1$ case. As can be seen from Fig. 3b, zero-energy end states do indeed occur for $m = 1, 2$ and 3, indicating that, according to the tight-binding calculations, the experimentally realized 7-AGNR-S(1,3) belongs to the topologically trivial $\mathbb{Z}_2 = 0$ class.

To verify this finding experimentally, the synthetic route shown in Fig. 2 was modified to allow the required extension of the staggered nanoribbon structure with a pristine 7-AGNR backbone segment. This is realized by sequential deposition of monomer **1** for the 7-AGNR-S(1,3) and dibromo-bianthryl (DBBA, monomer **2**) for the 7-AGNR (Fig. 3c, Supplementary Fig. 25). Differential conductance

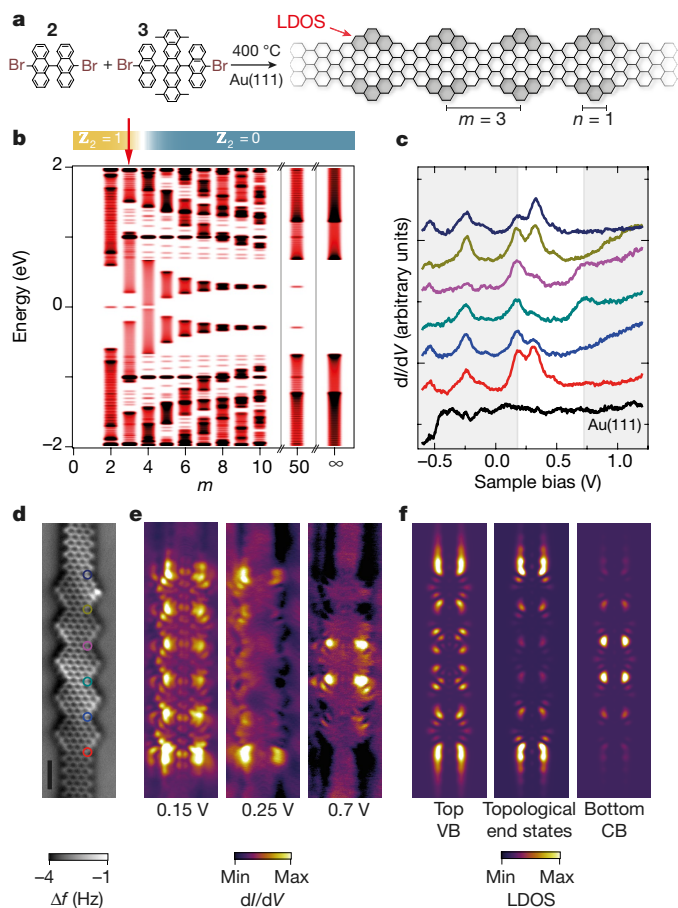


Fig. 4 | Non-trivial topological ($\mathbb{Z}_2 = 1$) phase of the inline edge-extended 7-AGNR-I(1,3) structure. **a**, On-surface synthesis route to the 7-AGNR backbone extended 7-AGNR-I(1,3) nanoribbon. **b**, LDOS plots evaluated at the end of the 7-AGNR-I(1, m) segment (see arrow in **a**) as a function of inter-segment spacing m , revealing the $\mathbb{Z}_2 = 1$ to $\mathbb{Z}_2 = 0$ transition at $m = 4$ with nearly complete gap closure and disappearance of the zero energy states for $m > 3$. **d**, Constant-height nc-AFM frequency shift (Δf) image of a 5-unit 7-AGNR-I(1,3) segment with 7-AGNR extensions at both ends. **c**, dI/dV spectra (-0.6 V and 100 pA set-point before opening feedback loop) taken at the locations indicated by the markers of corresponding colour in **d**. **e**, Experimental dI/dV maps of the main spectroscopic features at $+0.15$ V, $+0.25$ V and $+0.7$ V (all with $I = 500$ pA). **f**, Tight-binding-simulated charge-density maps at the top of the VB, at $E = 0$ eV, and at the bottom of the CB, computed for the experimental structure (**d**). The 1 nm scale bar in **d** applies also to panel **e** and **f**.

dI/dV spectroscopy at the end of the SSH GNR segment (red curve and marker in Fig. 3d and e) and at the internal SSH chain site (blue) shows nearly identical spectra with no indication of an end state. This is further corroborated by dI/dV mapping at selected energies around $E = 0$ eV (Fig. 3f). At $U = -0.05$ V the onset of the spatially extended VB states of the 7-AGNR-S(1,3) can be seen; $U = 0.25$ V corresponds to a gap with no particular features, and at $+0.65$ V the bottom of the CB can be observed, in good agreement with the tight-binding charge-density simulation of the lowest-energy CB state of the 7-AGNR-7-AGNR-S(1,3) heterostructure (Fig. 3g). The experiment therefore confirms the tight-binding prediction that the 7-AGNR-S(1,3) is topologically trivial with $\mathbb{Z}_2 = 0$.

The staggered N -AGNR-S(n, m) exhibits boundary states only for nanoribbon widths $N = 3p + 1$ (where p is an integer) and provides an electronic *cis*-polyacetylene analogue. If instead of an asymmetric N -AGNR to $(N + 2)$ -AGNR junction an axially symmetric N -AGNR to $(N + 4)$ -AGNR junction is considered, as illustrated in Fig. 4a, the resulting 'in-line' edge-extended GNR will yield zero-energy boundary states for all backbone widths N (Supplementary Fig. 6). Similar to the

staggered structure, the $(N + 4)$ segment length is denoted by n and the segment spacing by m , and the in-line 'I' structure is thus labelled N -AGNR-I(n, m). The structure shown in Fig. 4a is therefore a 7-AGNR-I(1,3). The LDOS series for the 7-AGNR-I(1, m) family in Fig. 4b reveals topological end states for $m = 2$ and $m = 3$ with the non-trivial to trivial phase transition between $m = 3$ ($\mathbb{Z}_2 = 1$) and $m = 4$ ($\mathbb{Z}_2 = 0$). The bulk band structure for the 7-AGNR-S(1,3) exhibits two *trans*-polyacetylene-like bands with $t_n = 0.45$ eV and $t_m = 0.65$ eV (Supplementary Fig. 11), which is very similar to 7-AGNR-S(1,3). Topological phase diagrams for the 7-AGNR-S(n, m) and 7-AGNR-I(n, m) structures are given in Supplementary Fig. 14 for $n \in [1, 9]$ and $m \in [1, 9]$.

The synthetic route to the 7-AGNR extended 7-AGNR-I(1,3) structure is analogous to the one for the staggered structure, but using 6,13-bis(10-bromoanthracen-9-yl)-1,4,8,11-tetramethylpentacene (BATMP, monomer 3) as the precursor molecule. Figure 4d presents the nc-AFM image of a 5-unit 7-AGNR-I(1,3) that is extended by 7-AGNR segments at both ends. dI/dV spectra recorded at the 7-AGNR/7-AGNR-I(1,3) junction (dark blue and red in Fig. 4c and d) reveal a state at approximately 0.25 V that is only present at the chain ends, as confirmed by dI/dV mapping (Fig. 4e). Comparison with tight-binding calculations (Fig. 4f) reveals that the extended state at 0.15 V can be assigned to the top of the VB, the 4-lobe state at $+0.7$ V in the centre of the chain to the CB minimum, and that the state at $+0.25$ V is indeed the expected topologically non-trivial bulk-boundary end state (see Supplementary Fig. 26 for a high-resolution dI/dV map). This state is not observed at exactly 0 V owing to charge doping by the substrate, which is well known to occur for low-bandgap GNRs on Au(111)^{29,30} and its non-mid-gap position might be due to substrate-dependent many-body energy renormalization³¹. All together, this analysis shows that, in contrast to the staggered trivial 7-AGNR-S(1,3), the inline edge-extended 7-AGNR-I(1,3) belongs to the topologically non-trivial $\mathbb{Z}_2 = 1$ class and hosts topological end states.

For our discussion we have chosen the topological invariant related to the winding number of the SSH model (\mathbb{Z}_2) as identifier of the topological class. Alternatively, the Zak phase of all occupied bands can also be used, yielding the topological invariant¹⁸ \mathbb{Z}'_2 , which is $\mathbb{Z}'_2 = 1 - \mathbb{Z}_2$ for the structures considered here (see Supplementary Fig. 13).

The presence of short zigzag edge segments in the structure families discussed here suggests the possibility of magnetic ordering³². For the 7-AGNR-S(1,3) and 7-AGNR-I(1,3) structures the relatively strong coupling suppresses magnetic ordering, but the formation of antiferromagnetic spin-chains is expected for structures with larger n and m (Supplementary Figs. 22–23). A more direct effect of the (n, m) -dependent coupling strength is that the bandgap can be tuned over a wide range without changing the ribbon width (Supplementary Fig. 14).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0375-9>.

Received: 1 February 2018; Accepted: 7 June 2018;

Published online 8 August 2018.

- Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
- König, M. et al. Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007).
- Kitaev, G. Unpaired Majorana fermions in quantum wires. *Phys. Usp.* **44**, 131–136 (2001).
- Bradlyn, B. et al. Topological quantum chemistry. *Nature* **547**, 298–305 (2017).
- de Vries, E. K. et al. Towards the understanding of the origin of charge-current-induced spin voltage signals in the topological insulator Bi₂Se₃. *Phys. Rev. B* **92**, 201102 (2015).
- Mourik, V. et al. Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* **336**, 1003–1007 (2012).

7. Chiang, C. K. et al. Electrical conductivity in doped polyacetylene. *Phys. Rev. Lett.* **39**, 1098 (1977).
 8. Su, W.-P., Schrieffer, J. R. & Heeger, A. J. Soliton excitations in polyacetylene. *Phys. Rev. B* **22**, 2099–2111 (1980).
 9. Longuet-Higgins, H. C. & Salem, F. R. S. L. The alternation of bond lengths in long conjugated chain molecules. *Proc. R. Soc. Lond. A* **251**, 172–183 (1959).
 10. Su, W., Schrieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698 (1979).
 11. Asbóth, J. K., Oroszlány, L. & Pályi, O. A short course on topological insulators: band-structure topology and edge states in one and two dimensions *Lecture Notes in Physics* Vol. 919 (Springer, Cham, 2016).
 12. Meier, E. J., An, F. A. & Gadway, B. Observation of the topological soliton state in the Su–Schrieffer–Heeger model. *Nat. Commun.* **7**, 13986 (2016).
 13. Tan, W., Sun, Y., Chen, H. & Shen, S.-Q. Photonic simulation of topological excitations in metamaterials. *Sci. Rep.* **4**, 3842 (2015).
 14. Chen, B. G.-g., Upadhyaya, N. & Vitelli, V. Nonlinear conduction via solitons in a topological mechanical insulator. *Proc. Natl Acad. Sci. USA* **111**, 13004–13009 (2014).
 15. Cai, J. et al. Atomically precise bottom-up fabrication of graphene nanoribbons. *Nature* **466**, 470–473 (2010).
 16. Nguyen, G. D. et al. Atomically precise graphene nanoribbon heterojunctions from a single molecular precursor. *Nat. Nanotechnol.* **12**, 1077–1082 (2017).
 17. Talirz, L., Ruffieux, P. & Fasel, R. On-surface synthesis of atomically precise graphene nanoribbons. *Adv. Mater.* **28**, 6222–6231 (2016).
 18. Cao, T., Zhao, F. & Louie, S. G. Topological phases in graphene nanoribbons: junction states, spin centers, and quantum spin chains. *Phys. Rev. Lett.* **119**, 076401 (2017).
 19. Wang, Z. et al. Strong interface-induced spin–orbit interaction in graphene on WS₂. *Nat. Commun.* **6**, 8339 (2015).
 20. Feigel'man, M. V., Skvortsov, M. A. & Tikhonov, K. S. Theory of proximity-induced superconductivity in graphene. *Solid State Commun.* **149**, 1101–1105 (2009).
 21. Nadj-Perge, S. et al. Observation of Majorana fermions in ferromagnetic atomic chains on a superconductor. *Science* **346**, 602–607 (2014).
 22. Shen, Q., Gao, H.-Y. & Fuchs, H. Frontiers of on-surface synthesis: from principles to applications. *Nano Today* **13**, 77–96 (2017).
 23. Fairbrother, A. et al. High vacuum synthesis and ambient stability of bottom-up graphene nanoribbons. *Nanoscale* **9**, 2785–2792 (2017).
 24. Llinas, J. P. et al. Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons. *Nat. Commun.* **8**, 633 (2017).
 25. Wakabayashi, K., Sasaki, K., Nakanishi, T. & Enoki, T. Electronic states of graphene nanoribbons and analytical solutions. *Sci. Technol. Adv. Mater.* **11**, 054504 (2010).
 26. Söde, H. et al. Electronic band dispersion of graphene nanoribbons via Fourier-transformed scanning tunneling spectroscopy. *Phys. Rev. B* **91**, 045429 (2015).
 27. Deniz, O. et al. Revealing the electronic structure of silicon intercalated armchair graphene nanoribbons by scanning tunneling spectroscopy. *Nano Lett.* **17**, 2197–2203 (2017).
 28. Wang, S. et al. Giant edge state splitting at atomically precise graphene zigzag edges. *Nat. Commun.* **7**, 11507 (2016).
 29. Kimouche, A. et al. Ultra-narrow metallic armchair graphene nanoribbons. *Nat. Commun.* **6**, 10177 (2015).
 30. Merino-Díez, N. et al. Width-dependent band gap in armchair graphene nanoribbons reveals Fermi level pinning on Au(111). *ACS Nano* **11**, 11661–11668 (2017).
 31. Kharche, N. & Meunier, V. Width and crystal orientation dependent band gap renormalization in substrate-supported graphene nanoribbons. *J. Phys. Chem. Lett.* **7**, 1526–1533 (2016).
 32. Fujita, M., Wakabayashi, K., Nakada, K. & Kusakabe, K. Peculiar localized state at zigzag graphite edge. *J. Phys. Soc. Jpn* **65**, 1920–1923 (1996).
- Acknowledgements** This work was supported by the Swiss National Science Foundation, the Office of Naval Research BRC Program, the European Union's Horizon 2020 research and innovation programme (GrapheneCore1 696656), and the NCCR MARVEL. C.A.P. thanks the Swiss Supercomputing Center (CSCS) for computational support. X.Y. is grateful for a fellowship from the China Scholarship Council. O.G. thanks O. Yazyev, D. Rizzo and D. Bercioux for discussions.
- Reviewer information** *Nature* thanks T. Heine, I. Swart and the other anonymous reviewer(s) for their contribution to the peer review of this work.
- Author contributions** O.G., P.R. and R.F. conceived and supervised this work. A.N., X.Y., X.F. and K.M. designed and synthesized the molecular precursors. S.W. performed the on-surface synthesis and scanning probe microscopy characterization. G.B.B. did the Raman analysis, C.D., A.C. and V.M. performed the corresponding simulations. C.A.P. did the DFT calculations. O.G. developed the conceptual framework, performed the tight-binding calculations and wrote the manuscript, with contributions from all co-authors. P.R., S.W. and O.G. designed the figures, with contributions from other co-authors.
- Competing interests** The authors declare no competing interests.
- Additional information**
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0375-9>.
Reprints and permissions information is available at <http://www.nature.com/reprints>.
Correspondence and requests for materials should be addressed to O.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Tight-binding calculations of electronic structure. The calculations of the electronic structure are based on the nearest-neighbour hopping tight-binding Hamiltonian considering the $2p_z$ orbital of the carbon atoms only:

$$H = \sum_i \varepsilon_i c_i^\dagger c_i - \sum_{\langle i,j \rangle} \gamma_0 c_i^\dagger c_j \quad (1)$$

Here c_i^\dagger and c_i denote the usual creation and annihilation operators on site i . $\langle i,j \rangle$ denotes the summation over nearest-neighbour sites, the on-site energies ε_i are all set to zero, and the nearest-neighbour hopping parameter is chosen to be $\gamma_0 = 3$ eV.

Band structures are calculated by taking into account the wave vector-dependent complex Bloch phase factors in the tight-binding Hamiltonian. Unfolding of the band structures into the extended Brillouin zone is achieved by projection of the wavefunctions of energy $E_n(\mathbf{k}_\parallel)$ on plane waves $|\langle \mathbf{k}_\parallel + \mathbf{k}_\perp | \psi_n(\mathbf{k}_\parallel) \rangle|$. The corresponding weight is displayed by marker size and colour. Here \mathbf{k}_\parallel and \mathbf{k}_\perp denote the wave vectors parallel and perpendicular to the GNR axis, respectively. The perpendicular wave vector \mathbf{k}_\perp for the projection is chosen to be non-zero in order to cut through the Dirac point of the parent graphene structure at $\mathbf{k}_\perp = 2\pi/3a$ and $\mathbf{k}_\parallel = 2\pi/\sqrt{3}a$ with $a = 2.44$ Å being the length of the graphene basis vector.

Wavefunctions are reconstructed from the tight-binding eigenvectors $\alpha_{i,n}$ of energy E_n by summing up the carbon $2p_z$ Slater-type orbitals with $\xi = 1.625$ atomic units over the atomic sites i of the structure.

$$\psi_n(\mathbf{r}) = \sum_i \alpha_{i,n} \exp(-\xi |\mathbf{r} - \mathbf{r}_i|) \quad (2)$$

STS-mapping simulations are achieved to a first approximation by displaying the charge density of the states considered in the energy interval $[\varepsilon_1, \varepsilon_2]$ at constant height z_0 according to:

$$\text{LDOS}(x, y, z_0) = \sum_n |\psi_n(\mathbf{r})|^2 \text{ for all } n \text{ with } E_n \in [\varepsilon_1, \varepsilon_2] \quad (3)$$

The results of the band-structure calculations for the 7-AGNR-S(1,3) and 7-AGNR-I(1,3) structures are compared to DFT calculations in Supplementary Fig. 12.

Molecular precursor and nanoribbon synthesis. The chemical synthesis of the monomers **1** (BADMT), **2** (BATMP) and **3** (DBBA) is detailed in the Supplementary Information together with details of the on-surface synthesis of the corresponding GNRs (Supplementary Scheme 1, Supplementary Figs. 28–38, 24–26).

STM/STS and nc-AFM characterization. A commercial low-temperature STM/AFM system (Scienta Omicron) with a base pressure below 1×10^{-10} mbar was used for sample preparation and characterization under ultrahigh-vacuum conditions. STM images and differential conductance dI/dV maps were recorded in constant-current mode unless noted otherwise. Constant-height tunnelling current and nc-AFM frequency shift images were recorded with a CO-functionalized tip attached to a quartz tuning fork sensor (resonance frequency 23.5 kHz). dI/dV spectra were recorded using the lock-in technique ($U_{\text{rms}} = 20$ mV at 680 Hz modulation). All data shown were acquired at a sample temperature of 5 K.

Data availability. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability. The tight-binding calculations were performed using a custom-made code on the WaveMetrics IGOR Pro platform. Details of this tight-binding code can be obtained from the corresponding author on reasonable request.

Diode fibres for fabric-based optical communications

Michael Rein^{1,2,3}, Valentine Dominique Favrod^{1,4}, Chong Hou^{2,3}, Tural Khudiyev^{2,3}, Alexander Stolyarov⁵, Jason Cox⁶, Chia-Chun Chung⁶, Chhea Chhav⁶, Marty Ellis⁷, John Joannopoulos^{2,3,8} & Yoel Fink^{1,2,3,6*}

Semiconductor diodes are basic building blocks of modern computation, communications and sensing¹. As such, incorporating them into textile-grade fibres can increase fabric capabilities and functions², to encompass, for example, fabric-based communications or physiological monitoring. However, processing challenges have so far precluded the realization of semiconducting diodes of high quality in thermally drawn fibres. Here we demonstrate a scalable thermal drawing process of electrically connected diode fibres. We begin by constructing a macroscopic preform that hosts discrete diodes internal to the structure alongside hollow channels through which conducting copper or tungsten wires are fed. As the preform is heated and drawn into a fibre, the conducting wires approach the diodes until they make electrical contact, resulting in hundreds of diodes connected in parallel inside a single fibre. Two types of in-fibre device are realized: light-emitting and photodetecting p–i–n diodes. An inter-device spacing smaller than 20 centimetres is achieved, as well as light collimation and focusing by a lens designed in the fibre cladding. Diode fibres maintain performance throughout ten machine-wash cycles, indicating the relevance of this approach to apparel applications. To demonstrate the utility of this approach, a three-megahertz bi-directional optical communication link is established between two fabrics containing receiver–emitter fibres. Finally, heart-rate measurements with the diodes indicate their potential for implementation in all-fabric physiological-status monitoring systems. Our approach provides a path to realizing ever more sophisticated functions in fibres, presenting the prospect of a fibre ‘Moore’s law’ analogue through the increase of device density and function in thermally drawn textile-ready fibres.

Efforts to increase fibre functions can lead to substantial advantages because the inherent scalability of textile production can be harnessed to produce functional fabrics at a large scale^{3,4}. The preform-to-fibre drawing process has been demonstrated to deliver considerable functional capabilities on the fibre and textile level through the incorporation of materials with disparate electronic and optical properties into monofilaments^{4–10}. Nevertheless, this process has been limited to materials that could be co-drawn¹⁰ in their viscous states and are typically inferior in performance to ‘device-grade’ materials that are made using wafer-based approaches^{11,12}. In this work, we combine scalable preform-to-fibre drawing with high-performance prefabricated semiconductor devices. Specifically, we incorporate functional semiconductor devices and electrical conductors into a polymer-clad preform, where the viscous polymer cladding simultaneously facilitates device packaging and electrical connectorization in situ during the thermal draw. This process enables new fibre and textile optical communication functionality at unprecedented data rates, as well as a viable path to introducing a gamut of alternative electronic devices into thermally drawn fibres.

The fabrication approach used to produce these fibres is illustrated in Fig. 1a, b. Prefabricated semiconductor devices are embedded in

prescribed locations along the preform. As the preform is thermally drawn, the diodes separate axially while their lateral position is preserved by the surrounding viscous polymer. During the drawing process, electrical conductors are unspooled into hollow channels flanking the diodes. The lateral separation of these wires is gradually reduced in the neck-down region until electrical contact is made with the devices. In contrast with previously reported work, where low-melting-temperature metals were thermally co-drawn in polymer fibres^{5,10,13}, this work demonstrates the ability to embed high-melting-temperature tungsten or copper metallic wires in the fibres during the draw, thus providing highly conductive electrical conduits for the devices. It is worth noting that neither the wires nor the diodes scale down in size during the draw, nor they are in contact with each other in the preform. The preform design and the drawing process itself facilitate the electrical connection between the wires and the devices, as shown in Fig. 1b, c.

Fibres produced using this approach result in a linear array of semiconductor devices uniformly spaced along the fibre length and electrically connected in parallel, with fibre size as small as 350 μm \times 350 μm . Electrical connection to the in-fibre electrodes is achieved by stripping away the polymer cladding at one end of the fibre. When a voltage is applied to the embedded wires, the in-fibre light-emitting diodes (LEDs) emit light, as shown in Fig. 2a for several fibre samples containing LEDs of different colour. This discovery is the first demonstration of a thermally drawn fibre with embedded semiconductor devices that are able to emit light when the fibre is supplied with electrical current, circumventing the necessity of applying external coatings or conductors^{14–20}. Moreover, unlike other^{17–20} approaches that yield short fibre lengths, the current approach enables kilometres of functional fibre to be drawn from a single preform with more than a hundred discrete devices connected in parallel throughout the entire fibre. The linear device density in the fibres can be directly controlled by varying the linear density of devices in the preform (for a given preform-to-fibre draw-down ratio) or by introducing several layers of devices and wires in the preform, as demonstrated in Extended Data Fig. 1. For example, for a draw-down ratio of 40, we are able to reduce the inter-diode distance from 2 m to approximately 17 cm. Addition of more layers in the preform will potentially lead to even higher device density in the fibres.

This technique is not limited to the incorporation of LEDs into fibres; other electronic devices could be embedded within thermally drawn fibres in a similar fashion. We embedded p–i–n photodetectors into fibres to enable high-bandwidth photodetection, in contrast to amorphous chalcogenide photoresistive materials previously drawn in fibres¹⁰, which have much lower responsivity and bandwidth compared to crystalline semiconductors such as Si, Ge or GaAs. The method used for the introduction of crystalline GaAs semiconducting p–i–n photodiodes is shown in Extended Data Fig. 2. Characterization of the photodetecting fibres was carried out by illuminating them with the red LED fibres. Figure 2e shows a clear rectifying behaviour of the

¹Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Research Laboratory of Electronics (RLE), Massachusetts Institute of Technology, Cambridge, MA, USA. ³Institute for Soldier Nanotechnologies, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Department of Materials Science, Swiss Institute of Technology (EPFL), Lausanne, Switzerland. ⁵MIT Lincoln Laboratory, Lexington, MA, USA. ⁶Advanced Functional Fabrics of America (AFFOA), Cambridge, MA, USA. ⁷Inman Mills, Inman, SC, USA. ⁸Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. *e-mail: yoel@mit.edu

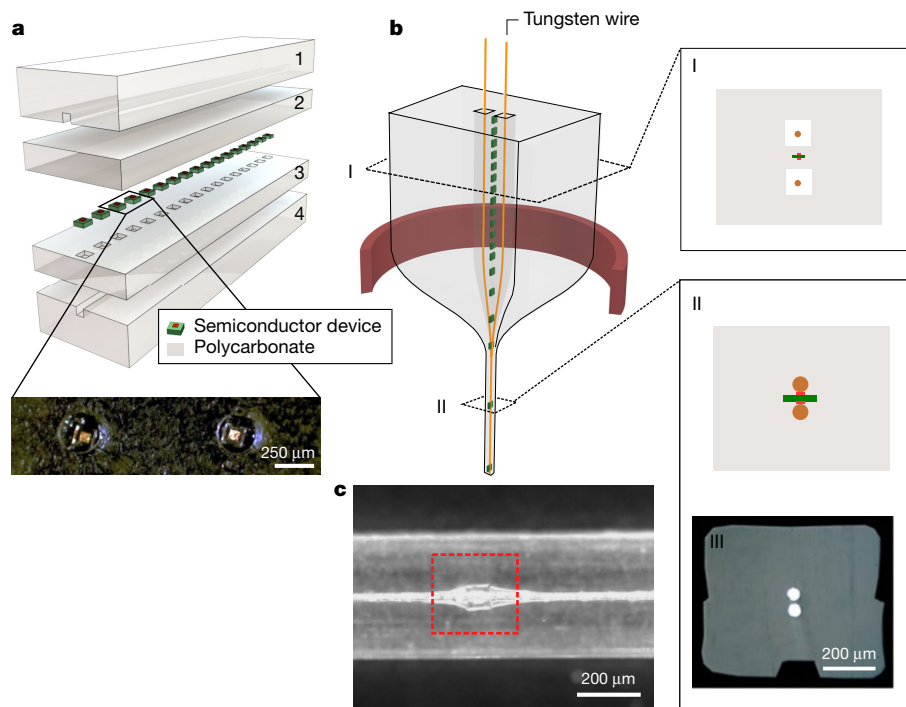


Fig. 1 | Preform structure and fibre drawing results. **a**, Illustration of the preform structure. It is composed of two main slabs (1 and 4), with a groove milled on their surfaces along the entire length of the preform to accommodate the metallic wires that are interfaced with the devices in the fibre. Numerous pockets with the size of the devices (of the order of $100\text{ }\mu\text{m}$) are milled in the inner layer (3) to accommodate the electronic devices, as shown in the inset. A top polymeric layer (2) is placed on top of the layer containing the devices, and the preform is thermally consolidated in a heated hydraulic press. **b**, Illustration of the preform drawing process. The metallic wires (orange) are fed through the preform, which is heated and drawn (red ring). The metallic wires and devices are then embedded

and packaged in the fibres. Inset I, illustration of the preform cross-section, showing the devices (green rectangle with red contact pads), and wires (orange circles) placed in the grooves (white rectangles). The grooves are larger than the wire in the preform. Inset II, illustration of the fibre cross-section. The devices and wires are well embedded in the fibre cladding, where the wires are touching the contact pads of the devices. Inset III, optical micrograph of the fibre cross-section, showing two tungsten wires embedded in the fibre cladding, without any visible gaps or electrical short-circuiting. **c**, Optical micrograph of the fibre (side view), showing a LED device (inside the dashed red square) and the wires in contact.

photodetecting fibre. In the reverse-bias regime, a substantial increase (about four orders of magnitude) in the photocurrent is observed under illumination compared to the dark current. The measured bandwidth of the photodetecting fibres is shown in Fig. 2f, where the 3-dB bandwidth is found to be 3 MHz—an improvement by orders of magnitude compared to photodetecting fibres based on chalcogenide semiconductors²¹. The limit to the measured bandwidth could be attributed to the parasitic capacitance between the long metallic wires in the fibre. Additionally, the amplifier of the measurement system limits the operation of the system at high frequencies, introducing a trade-off between measured signal strength and system speed.

High-speed fibre LED transmitters and photodetectors present an opportunity for high-bandwidth inter-fibre communication links. Moreover, inter-textile communications functionality is achieved because all components of the fibres are internal to the fibre structure and thus are able to withstand the strains and stresses of textile manufacturing techniques, such as weaving, and even day-to-day handling, including water immersion and machine washing, as shown in Extended Data Figs. 3, 4. Both LED and photodetecting fibres were woven into a separate textile polyester fabric using a conventional industrial loom in a satin weave pattern, as shown in Fig. 3a. Electrical connection to the fibres was made post-weaving at the fabric edge, and the fibres were found to be fully operational. In all cases, the fibres had identical performance as before weaving or the washing cycle tests, presenting a viable path towards the everyday use of this nascent capability.

The opportunity to control the cross-section of the fibre cladding adds an additional degree of freedom to enhance the performance of the fibres. The fibre cross-section can be designed as a lens, with the aim to increase the communication range by collimating and focusing

the light emitted and collected by the LED and photodetecting fibres, respectively. Analytic and numerical simulations were carried out to determine the optimal location of the devices in the fibre, as well as the shape of the cladding, to achieve maximal communication range (see Extended Data Figs. 5–9). This fibre cladding shape is shown in Fig. 3b, c. Experimental results on the effect of the lens cross-section on the communication range are shown in Fig. 3c, where the advantage is apparent for the lensed-cladding fibres compared to fibres with square cross-sections. Both curves follow the inverse-square intensity decay law, as discussed in Methods; nevertheless, shaping the fibre cladding as a collimating lens results in a higher optical flux at the plane of the photodetecting fibre, thus increasing the measured intensity at a given distance.

Integration of high-speed optical transmitters and receivers into fabrics present many compelling applications. First, we demonstrate a fabric-to-fabric communication scheme for two fabrics separated by 1 m in free space, as shown in Fig. 4a, b. Figure 4b shows the signal recorded by the photodetecting fibre when the LED fibre is driven with a frequency of 20 kHz, close to the maximum of the audible frequency range, demonstrating the ability to transmit audio signal over fabric. This capability could be exploited for numerous other applications, from fabric-enabled light fidelity technology²¹, to fabric-encrypted local information transfer and indoor positioning platforms. Second, we demonstrate the capabilities of these fibres in the context of physiological measurements. Specifically, we demonstrate a textile-based photoplethysmography system for pulse measurement²² based on the developed textile platform. Here, a green LED fibre is embedded in a cotton fabric sock adjacent to a GaAs photodetecting fibre, as illustrated in Fig. 4c. A pulse measurement is obtained by placing an index finger on both fibres. The change of the

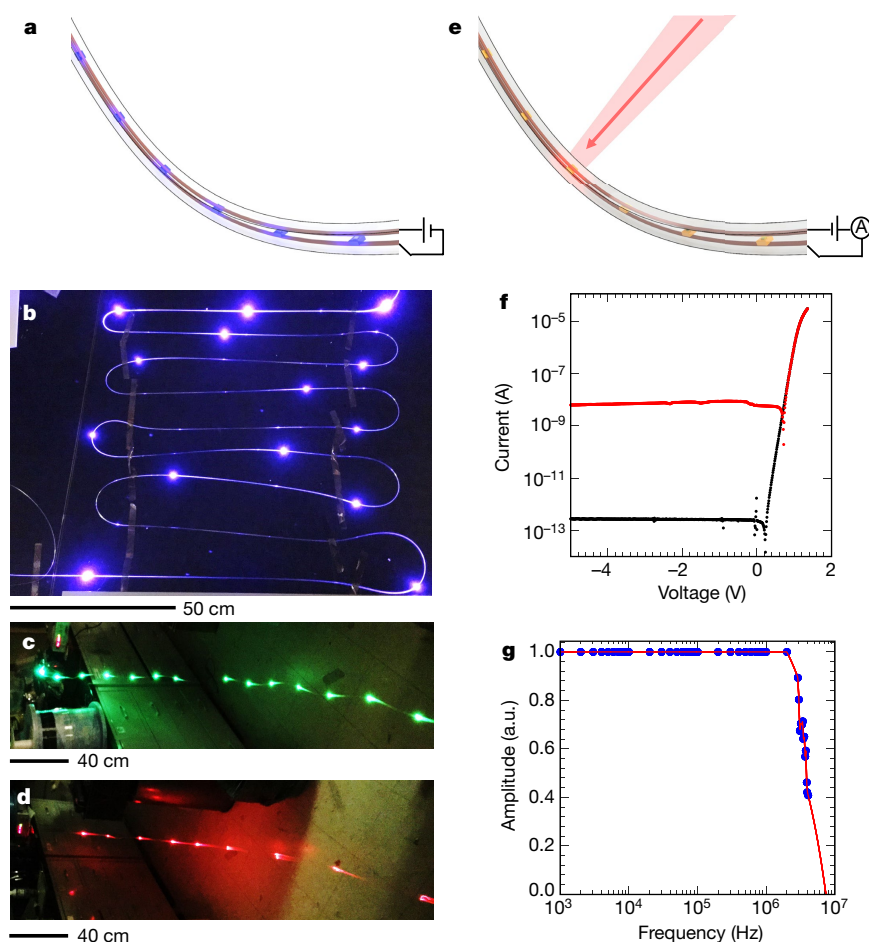


Fig. 2 | Light-emitting and high-bandwidth photodetecting fibres. **a–d**, Multimaterial fibres with light-emitting functionality. **a**, Illustration of the light-emitting fibres. The wires (orange) are connected to the LEDs (purple) and to a current supply (black lines) at the fibre end. **b**, Photograph of light-emitting fibres containing InGaN blue-colour LEDs. The devices appear every 370 ± 110 mm and the fibres are laid flat on a table. **c**, Fibres containing InGaN LEDs emitting green colour. The fibres were unspooled and held on the other end (right-hand side of the photograph). **d**, Fibres containing AlGaAsP LEDs emitting red colour. The fibres were unspooled and held on the other end (right-hand side of the photograph). **e–g**, High-bandwidth photodetecting fibres. **e**, Illustration of the photodetecting fibre structure, where an individual photodiode (orange) interacts with an external beam of light (red arrow). **f**, Current–voltage curve of a fibre containing one GaAs device, showing a clear rectifying behaviour. The black curve was obtained in darkness and the red curve under illumination. In the reverse-bias regime the current increases by a few orders of magnitude when the fibre is illuminated. The plot shows the absolute value of the current on a logarithmic scale. Application of a logarithmic function on low voltage and current values shows a kink in the response, which is not present in the raw data. **g**, Bandwidth measurement (blue circles) of the photodetecting fibre. The 3 dB bandwidth achieved is around 3 MHz. a.u., arbitrary units. The red line is a guide for the eye.

measured light intensity recorded by the photodetecting fibre due to the change in the light reflectance from the skin is shown in Fig. 4d. The measured signal directly correlates with volume changes in small blood vessels, which expand and constrict with every heartbeat. These results demonstrate the potential to integrate physiological sensors fully within fibres and textiles, not as add-ons to fabrics. The results presented here demonstrate a new paradigm for integrating pre-fabricated high-performance semiconductor devices into a fibre

form factor, paving the way towards increasingly functional fibre and fabric systems. We envision that this technology will enable new technological advances in the textile and apparel domains, telecommunications, as well as in biological and medical sciences. In particular, multifunctional fibres could enable a new generation of optogenetically modified neuron fibre probes²³, active media for textile–bacteria interaction²⁴ systems or active textiles with fragrance- or medicine-release capabilities²⁵.

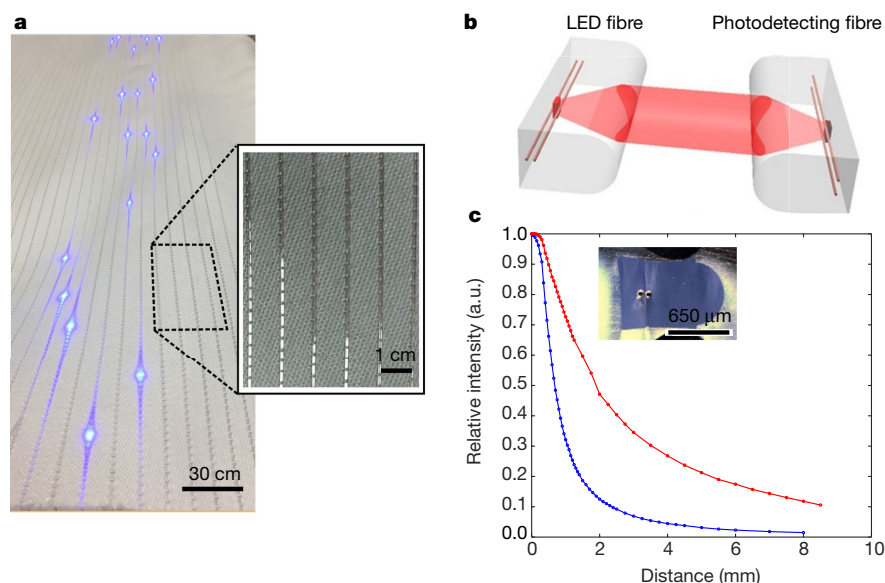


Fig. 3 | Embedding of fibres in fabrics and light collimation by the fibre cladding.

a, Light-emitting and photodetecting fibres embedded in a fabric. The blue-colour light-emitting fibres are embedded in a fabric and in operation. Inset, a closer look at the fibre–fabric interface. **b**, Communication between two lensed fibres. Left fibre, light-emitting fibre; right fibre, photodetecting fibre; red-shaded area, emitted light. The fibre lens could be considered as a cylindrical lens extending along the whole length of the fibre. **c**, Measured current of the photodetecting fibre, normalized with respect to the current measured at a contact between fibres, as a function of the distance between fibres. Blue symbols, no collimation or focusing; red symbols, light collimation and light focusing on the photodiode; blue and red lines, guides for the eye. Inset, optical micrograph of the light-emitting fibre, showing the lens and the two tungsten wires embedded in the cladding.

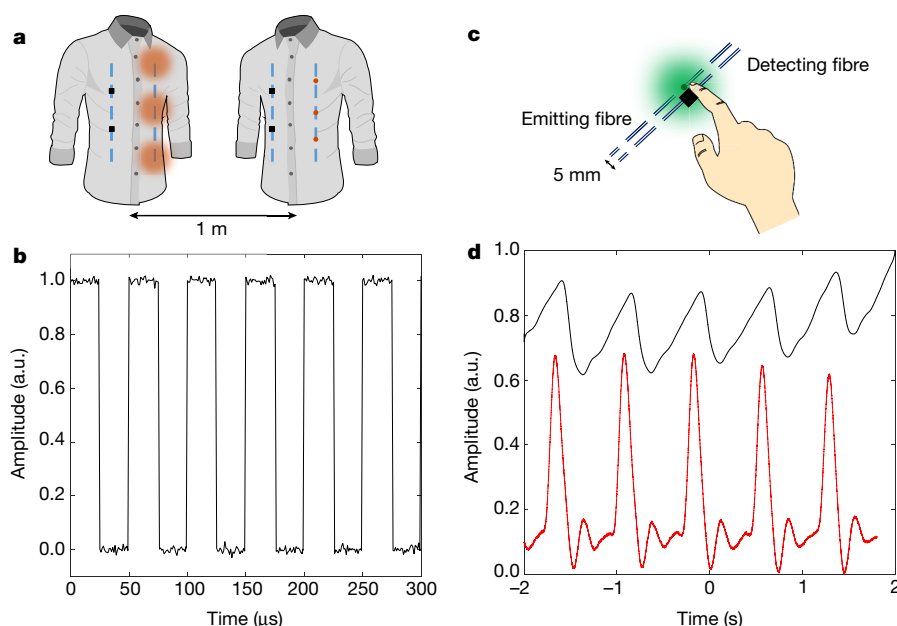


Fig. 4 | Applications of light-emitting and light-detecting fabric. **a**, Illustration of bi-directional communication system concept. A garment is composed of a fabric that contains both light-emitting (light blue dashed line with red circles) and photodetecting (light blue dashed line with black squares) fibres. The light-emitting fibres are modulated to transmit information that is being recorded by the photodetecting fibres in the other garment, placed at a distance of 1 m from each other. **b**, Experimental results of the current recorded by the photodetecting fibres incorporated into a fabric. The light was emitted from LED fibres embedded in another fabric located at a distance of 1 m from the photodetecting fabric. The light-emitting fibres were

connected to a function generator delivering a square-wave signal with a frequency of 20 kHz. The transmission of the signal was recorded by the fibres (see Methods for additional details). **c**, Illustration of a photoplethysmography pulse measurement setup using light-emitting (dashed line with green light) and photodetecting (dashed line with black square) fibres placed at a distance of 5 mm from each other. Placing a finger on both fibres allows recording the reflected light, which is sensitive to blood circulation in the blood vessels close to the skin. **d**, Experimental results of the current measured by the photodetecting fibre (black curve) compared to the output of a commercial pulse sensor (red curve). Periodic changes in the recorded intensity correspond to the frequency of the pulse.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0390-x>.

Received: 4 January 2018; Accepted: 8 June 2018;

Published online 8 August 2018.

- Kitai, A. in *Principles of Solar Cells, LEDs and Diodes: The Role of the PN Junction* Chs. 2, 5 (John Wiley & Sons, Chichester, 2011).
- Stoppa, M. & Chiolerio, A. Wearable electronics and smart textiles: a critical review. *Sensors* **14**, 11957–11992 (2014).
- Abouraddy, A. F. et al. Towards multifunctional multifunctional fibres that see, hear, sense and communicate. *Nat. Mater.* **6**, 336–347 (2007).
- Tao, G., Abouraddy, A. F., Stolyarov, A. M. & Fink, Y. in *Lab-on-Fiber Technology* (eds Cusano, A. et al.) 1–26 (Springer International Publishing Switzerland, 2015).
- Gumennik, A. et al. All-in-fiber chemical sensing. *Adv. Mater.* **24**, 6005–6009 (2012).
- Lestoquoy, G., Chocat, N., Wang, Z., Joannopoulos, J. D. & Fink, Y. Fabrication and characterization of thermally drawn fiber capacitors. *Appl. Phys. Lett.* **102**, 152908 (2013).
- Hart, S. D. et al. External reflection from omnidirectional dielectric mirror fibers. *Science* **296**, 510–513 (2002).
- Chocat, N. et al. Piezoelectric fibers for conformal acoustics. *Adv. Mater.* **24**, 5327–5332 (2012).
- Egusa, S. et al. Multifunctional piezoelectric fibres. *Nat. Mater.* **9**, 643–648 (2010).
- Bayindir, M. et al. Metal-insulator-semiconductor optoelectronic fibres. *Nature* **431**, 826–829 (2004).
- Rein, M. et al. Self-assembled fibre optoelectronics with discrete translational symmetry. *Nat. Commun.* **7**, 12807 (2016).
- Borisova, Z. U. *Glassy Semiconductors* (Plenum Press, New York, 1981).
- Sorin, F., Lestoquoy, G., Danto, S., Joannopoulos, J. D. & Fink, Y. Resolving optical illumination distributions along an axially symmetric photodetecting fiber. *Opt. Express* **18**, 24264–24275 (2010).
- Cherenack, K. H., Kinkeldei, T., Zysset, C. & Tröster, G. Woven thin-film metal interconnects. *IEEE Electron Device Lett.* **31**, 740–742 (2010).
- Cherenack, K., Zysset, C., Kinkeldei, T., Münzenrieder, N. & Tröster, G. Woven electronic fibers with sensing and display functions for smart textiles. *Adv. Mater.* **22**, 5178–5182 (2010).
- Dias, T. & Ratnayake, A. in *Electronic Textiles* 109–116 (Woodhead Publishing, Cambridge, 2015).
- Kwon, S. et al. High luminance fiber-based polymer light-emitting devices by a dip-coating method. *Adv. Electron. Mater.* **1**, 1500103 (2015).
- Zhang, Z. et al. A colour-tunable, weavable fibre-shaped polymer light-emitting electrochemical cell. *Nat. Photon.* **9**, 233–238 (2015).
- O'Connor, B., An, K. H., Zhao, Y., Pipe, K. P. & Shtein, M. Fiber shaped light emitting device. *Adv. Mater.* **19**, 3897–3900 (2007).
- O'Connor, B., Pipe, K. P. & Shtein, M. Fiber based organic photovoltaic devices. *Appl. Phys. Lett.* **92**, 2006–2009 (2008).
- Dimitrov, S. & Haas, H. *Principles of LED Light Communications* (Cambridge University Press, Cambridge, 2015).
- Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **28**, R1–R39 (2007).
- Canales, A. et al. Multifunctional fibers for simultaneous optical, electrical and chemical interrogation of neural circuits *in vivo*. *Nat. Biotechnol.* **33**, 277–284 (2015).
- Levskaia, A. et al. Engineering *Escherichia coli* to see light. *Nature* **438**, 441–442 (2005).
- Broichhagen, J. et al. Optical control of insulin release using a photoswitchable sulfonyleurea. *Nat. Commun.* **5**, 5116 (2014).

Acknowledgements This work was supported in part by the MIT Materials Research Science and Engineering Center (MRSEC) through the MRSEC Program of the National Science Foundation under award number DMR-1419807 and in part by the US Army Research Laboratory and the US Army Research Office through the Institute for Soldier Nanotechnologies, under contract number W911NF-13-D-0001, with funding provided by the Air Force Medical Services. This work was also supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract numbers FA8721-05-C-0002 and FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering. The authors express their gratitude to S. Maayani for discussions and simulations of the lensed fibre system; to D. Bono and C. Marcus for advice and support in building the fibre-based pulse measurement setup; to R. Yuan for illustration of the results presented in the manuscript; and to E. Simhon for discussions from research ideation through to its completion.

Reviewer information Nature thanks D. Richardson, M. Schmidt, M. Shtein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions M.R. and Y.F. conceived the study. M.R. designed the experiments, the preform structure and diode integration methodology, drew the fibres and characterized the fibres. V.D.F. designed the pulse measurement setup, performed the experiments and interpreted the results. C.H. and T.K. established the wire convergence system during the thermal draw of fibres. M.R., J.C., C.C.C. and C.C. drew fibres with Si photodetectors and built the electric circuitry for fabric-to-fabric communications. M.E. weaved the functional fibres in fabrics. M.R., A.S., J.J. and Y.F. reviewed and interpreted the results. M.R. and A.S. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0390-x>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Y.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Preform fabrication. Two slabs of polycarbonate (PC) (McMaster Carr #1749K149) were milled to introduce a trench running the entire length of the preform (~8 inches; 1 inch = 2.54 cm) with a width of 1.25 mm and depth of 1.6 mm. Two Teflon bars were inserted into milled pockets with similar size to prevent pocket collapse during preform consolidation. A thin, 500- μm -thick PC layer was consolidated on one of the PC slabs in a hydraulic hot press heated to a temperature of 175 °C for 5 min and then water-cooled. Small pockets were drilled in the thin PC layer to accommodate the microelectronic devices. The size of the pockets was slightly larger than that of the devices. The distance between the pockets was varied according to the desired density of the devices in the fibre. The devices were transferred manually to the drilled pockets in the preform while keeping the orientation of each device constant. Multiple devices were successfully integrated, such as LEDs (InGaN blue-colour LEDs, Cree C460UT170-0014-31; InGaN green-colour LEDs, Cree C527UT170-0108-31; AlGaInP red-colour LEDs, Three Five Materials TCO-07UOR), or various photodetectors (GaAs p-i-n photodiodes, Broadcom SPD2010; Si photodiodes, Three Five Materials PD-30027A-B). Another thin, 0.5-mm-thick layer was consolidated on top of the diodes to hold them in place in the preform. The top PC slab was consolidated on top of the layers to form the full preform. The final consolidation was performed in a hydraulic hot press at a temperature of 175 °C for an hour, and then the preform was slowly cooled to room temperature.

Fibre drawing. The fibres were fabricated by the thermal drawing process by placing the preform in a three-zone heating furnace, where the top, middle and bottom zones were heated to 150 °C, 270 °C and 110 °C, respectively. The preform was fed into the furnace at a rate of 1 mm min⁻¹ and drawn at a speed of 1.6 m min⁻¹, which resulted in a draw-down ratio of 40. Multiple tungsten (Goodfellow #343-809-07) or copper (Goodfellow #271-974-11) wires with a diameter of 50 μm were continuously fed into the preform during the draw. The pockets that accommodated the diodes and the wires in the fibre were smaller than the non-melting components, enabling full encapsulation of both the devices and the wires in the polymeric cladding; they also induced incision into the thin PC layers, which resulted in electrical connection between the wires and the terminals of the electronic devices. Away from the electronic devices, the PC layer was still present, and no short-circuiting occurred between the wires. Hundreds of metres of fibre was collected from each draw.

Optical microscopy characterization. To obtain the cross-section micrographs, the fibres were placed in a plastic holder (Struers Multiclips) and encapsulated in an epoxy matrix (Struers EpoFix), which was subsequently polished. The optical micrographs were obtained using a stereoscope microscope (Nikon SMZ745T).

Occasionally, some small air pockets could be observed around the devices after the draw. In most cases, these pockets did not interfere with the operation of the devices, as they were anchored in the surrounding polymeric cladding.

Operation of light-emitting fibres. The electrical wires in the fibre were exposed from the cladding by cutting the soft PC cladding. The exposed tungsten or copper wires were connected to a diode driver (Thorlabs LDC205C) and the current was supplied by the instrument, up to 30 mA for each light-emitting device.

Crystalline photodetectors characterization. To obtain the IV characteristics, the photodetecting fibres were connected to a power supply and pico-ammeter measurement system (Keithley 6487/6517A). The measurement was made in the dark and under illumination by a red light-emitting fibre placed at a distance of 10 mm from the photodetecting fibre. The operational bandwidth of the fibre photodetecting devices was measured using a function generator (Tektronix AFG3252) connected to a fibre pigtailed laser diode (Thorlabs LPM-660-SMA), configured as the illumination source. The electrical conductors of the photodetecting fibre were connected to a trans-impedance amplifier (Thorlabs TIA60) and to an oscilloscope (Agilent Technologies DSOX - 3014A). The frequency of the laser diode illumination swept a range of frequencies as the amplitude of the photodiode device voltage was measured with the oscilloscope at each frequency point.

Weaving fibres into textiles. The weaving of the fabric was carried out using a Picanol Gamma rapier weaving machine. In the warp direction, a satin weave design was used, and the yarn component was a blend of conventional nylon and cotton with a density of 100 threads per inch. In the weft direction, conventional filament polyester was used between the functional fibres at a density of 35 threads per inch. Device fibres were introduced into the fabric only in the weft direction. Although we did not measure the tensile force acting on the fibres during the process, later studies showed that the fibres can withstand tensile stress up to 70 MPa. We assume that the fibres should survive the weaving process as long as the tensile stress is kept under 70 MPa.

Information transmission from fabric to fabric. Red-light-emitting fibres were woven into a fabric, while Si photodetecting fibres were woven into a second fabric. The light-emitting fibres were connected to a function generator, and the photodetecting fibres were connected to a custom-built trans-impedance circuit, with the output of the circuit connected to an oscilloscope. The light-emitting fibre was

driven by a square wave (alternating voltage of 5 V) with a frequency of 20 kHz that electrically drove three devices per fibre, emitting a total optical power of 90 mW. The photodetecting fabric (with one photodetector embedded in each fibre) was placed at a distance of 1 m from the light-emitting fabric and obtained a signal that was recorded.

Photoplethysmography measurements. A green-light-emitting fibre was placed at a distance of 5 mm from a photodetecting fibre containing a GaAs detector. The light-emitting fibre was connected to a continuous current supply (Thorlabs LDC205C) and driven with a current of 20 mA. The photodetecting fibre was connected to a low-pass filter and the pulse measurement trace was collected with a computerized oscilloscope (Clevscope CS328A) when a finger was placed on both fibres.

Device density control in the fibres. The linear device density in the fibres can be directly controlled by varying their linear density in the preform ($\Delta l_{\text{preform}}$) and by varying the draw-down ratio (β), as demonstrated in Extended Data Fig. 1. The draw-down ratio is a ratio between the diameter of the preform (d_{preform}) and the diameter of the resulting fibre (d_{fibre}) and is described by the relation $\beta = d_{\text{preform}}/d_{\text{fibre}} = \sqrt{v_{\text{draw}}/v_{\text{feed}}}$, where v_{draw} is the fibre drawing speed and v_{feed} is the preform feed speed. According to the law of mass preservation, the axial distance in the preform will be translated to an axial distance in the fibre according to the relation $\Delta l_{\text{fibre}} = \beta^2 \Delta l_{\text{preform}}$. Thus, to decrease the inter-diode distance, we can vary the distance between the diodes in the preform or decrease the draw-down ratio.

For example, for a draw-down ratio of 40 and a linear device separation of 1.25 mm in the preform, the diodes in the fibre appear every $2,000 \pm 110$ mm. To increase the device density in the fibres, we can place them adjacent to each other in the preform, as demonstrated in inset (ii) of Extended Data Fig. 1b, and obtain a fibre with an inter-diode spacing of 370 ± 100 mm. The device dimensions set an upper limit on the highest linear device density that can be achieved in a fibre when the devices are placed in a straight line in the preform for a given draw-down ratio. To further increase the device density, these could be placed in several layers in the preform. This concept was demonstrated for two layers stacked vertically with a common anode wire, as shown in Extended Data Fig. 1c. The draw of this preform with the same draw-down ratio of 40 yields a fibre with an inter-diode separation of 173 ± 92 mm, increasing the effective linear density of the devices by a factor of two compared to a single device layer. Addition of more layers vertically or horizontally in the preform will potentially lead to a higher device density in the fibres. The measured dispersion in the device location is due to the finite precision of the positioning of the devices in the preform, which does not change when increasing the device density in the preform.

Extended Data Fig. 1d, e presents the effect of the inter-diode distance in the preform and the draw ratio on the distance between devices in the fibres, respectively. The effect of increasing the number of diode layers in the preform is demonstrated as well.

Increasing the communication range. An experiment was carried out to determine the dependence of the signal strength transmitted between a light-emitting fibre and a photodetecting fibre on the distance between the two fibres, as demonstrated in Extended Data Fig. 5. The figure shows that the recorded photocurrent is inversely proportional to the distance squared for distances larger than 1 mm. For shorter distances, the intensity decay diverges from this dependence owing to the finite size of the emitter and detector as well as the Lambertian radiation pattern of the LED. This intensity decay will affect the maximal communication distance, mostly because any signal below 0.1 nA is comparable to the ambient background noise in the experimental environment. This requires amplifying the signal. To increase the communication range substantially, a few approaches could be undertaken: (1) Increase the emitter intensity or reduce the beam divergence; (2) increase the receiver aperture size and photoelectric responsivity; (3) use appropriate electronic circuitry to measure the current output from the fibres.

Unfortunately, a larger device size means a larger fibre size—this is not desired because it makes the fibre stiffer, complicating the subsequent weaving process to integrate the fibres into fabrics. Higher illumination intensities lead to higher heat released from the semiconducting devices, causing them to heat to elevated temperatures and eventually melt the fibre cladding that encapsulates them. The electrical circuitry connected to the photodetecting fibre also influences the communication range. Amplifying the signal with high-gain circuitry will increase the communication range at the expense of noise amplification and operational bandwidth of the system. Alternatively, to increase the range of communication, the external shape of the fibre cladding could be optimized to collimate the light emitted from the light-emitting fibre and to focus the light on the photodetecting fibre, effectively introducing a cylindrical lens along the fibre length.

A photodetecting fibre was placed in front of a light-emitting fibre while the photocurrent from the photodetecting fibre was recorded as a function of the distance between the fibres. The results are shown in Extended Data Fig. 5, which demonstrates an inverse-distance-squared dependence at distances larger than

1 mm from the fibres, as expected from theory. Light collimation and focusing are expected to extend the communication range between the fibres. To find the optimal location of the devices in the fibre and the optimal shape of the cladding, a two-dimensional ray-optics numerical simulation was carried out. The size of the fibres was set to $500\text{ }\mu\text{m} \times 500\text{ }\mu\text{m}$. One fibre contained a photodetector (red rectangle, of size $250 \times 250\text{ }\mu\text{m}$) and the other fibre contained a finite-size source that emulated the LED. This source emitted light in a hemispherical configuration, with a Lambertian radiation pattern. The simulation was carried out using the ray-tracing module of the COMSOL software. This simulation was used to determine the optimal device location in the fibre and the optimal fibre shape. Adding a curved structure to the fibre surface and placing the device in the focal point of this lens is an effective approach to collimating and focusing the light between the light-emitting and photodetecting fibres. On the basis of lens physics, we should aim to place the devices in the focal point of the lens to both focus an external light source and to collimate the light emitted from the LEDs in the fibre. For a thick lens, the focal distance is given by the lens-maker's equation²⁶:

$$\frac{1}{f} = (n-1) \left(\frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)t}{nR_1R_2} \right) \quad (1)$$

Here, f is the focal length of the lens, n is the refractive index of the material of the lens, R_1 and R_2 are the radii of curvature of the two spherical parts of the lens and t is the lens thickness. Because the devices are embedded in the cladding, we have a curved surface only on the external side of the fibre, whereas the other side is flat, that is, with infinite radius. Thus, the focal point of such a lens will be as given by:

$$\frac{1}{f} = \frac{n-1}{R_1} \quad (2)$$

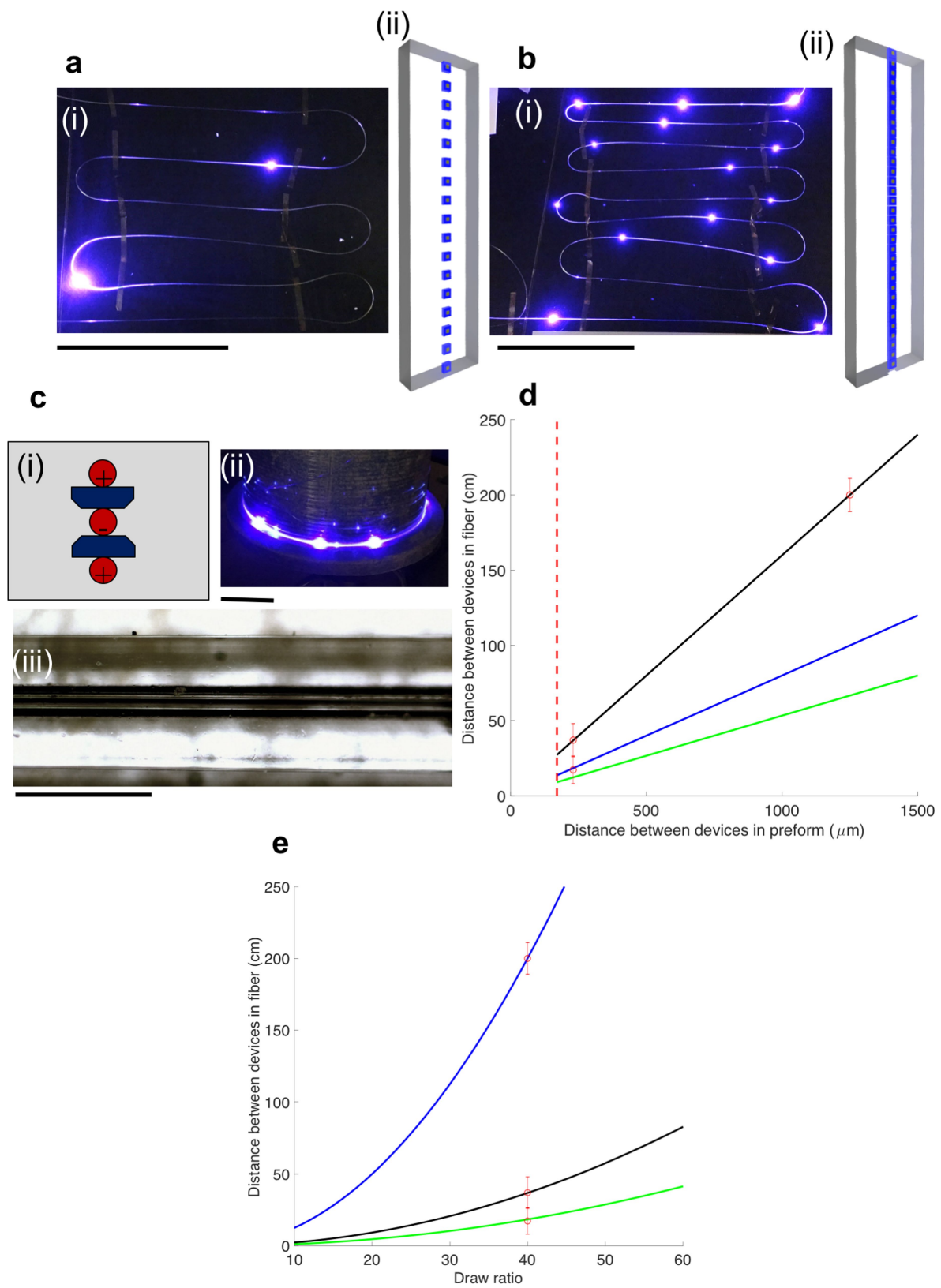
The cladding of the fibres is made of PC, which has an index of refraction of around $n = 1.58$ in the visible-wavelength domain. This locates the focal point of the fibre at a distance of $1.72R_1$ and defines the desired fibre structure, as shown in Extended Data Fig. 6a. Here we can see that the optimal structure is a rectangular or square fibre, where the device is located approximately $190\text{ }\mu\text{m}$ from the square's centre. The square is assumed to have side length of $500\text{ }\mu\text{m}$, equal to the lens diameter. We performed a ray-optics simulation to find the optimal location of the device in the fibre, and the ray-tracing results are shown in Extended Data Fig. 6b.

In this figure, we can see that some of the light is collimated by the lens, whereas some light is reflected or not collected by the lens and is allowed to escape the fibre without collimation. Other focusing techniques were considered; for example, using mirrors or multiple lenses on the surface of the fibre. Unfortunately, these geometries are harder to achieve in a fibre, and could be explored in the future. A similar approach was used to find the optimal device location and fibre shape for focusing a collimated beam of light on the photodetector. Extended Data Fig. 7 shows the results of a simulation carried out to determine the optimal location of the device in the fibre for collecting external light. Extended Data Fig. 7a shows the structure of the fibre and Extended Data Fig. 7b, c shows the results of the ray-tracing simulation and the dependence of the intensity on the location of the device in the fibre. We carried out a similar simulation for both types of fibres, one in front of the other, as shown in Extended Data Fig. 8. We can see that adding lenses on both fibres collimates and focuses some of the light, which will potentially extend the communication distance. We have reduced the optimized fibre structure to practice, as shown in Fig. 3c (inset).

Fibre device yield. Multiple optimization steps on the preform structure and drawing process have resulted in a yield of up to 95% (number of LEDs that light up relative to the total number of devices embedded in the preform). The main failure mechanisms observed in the process are misalignment between the metallic wires and the embedded devices, fibre polymer cladding or wire breakage during fibre drawing, lack of contact between the wires and the device contact pads, or occasional short-circuiting between the metallic wires in the fibres. Characterization results of the light-emitting fibres (measurement of the distance between adjacent diodes and of the emitted power) are shown in Extended Data Fig. 9. This characterization was performed on a fibre drawn from a preform containing two parallel rows of diodes, drawn down by a factor of 33. Most of the diodes appear at the expected distance. The power of each diode was measured and normalized with the maximum power emitted by the diode located closest to the driving circuit, which applied a constant voltage of 6 V. The emitted power decreases with the distance from the voltage source owing to the finite wire resistance. We drew more than 30 preforms with various diodes and all our draws yielded sections of working fibres.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

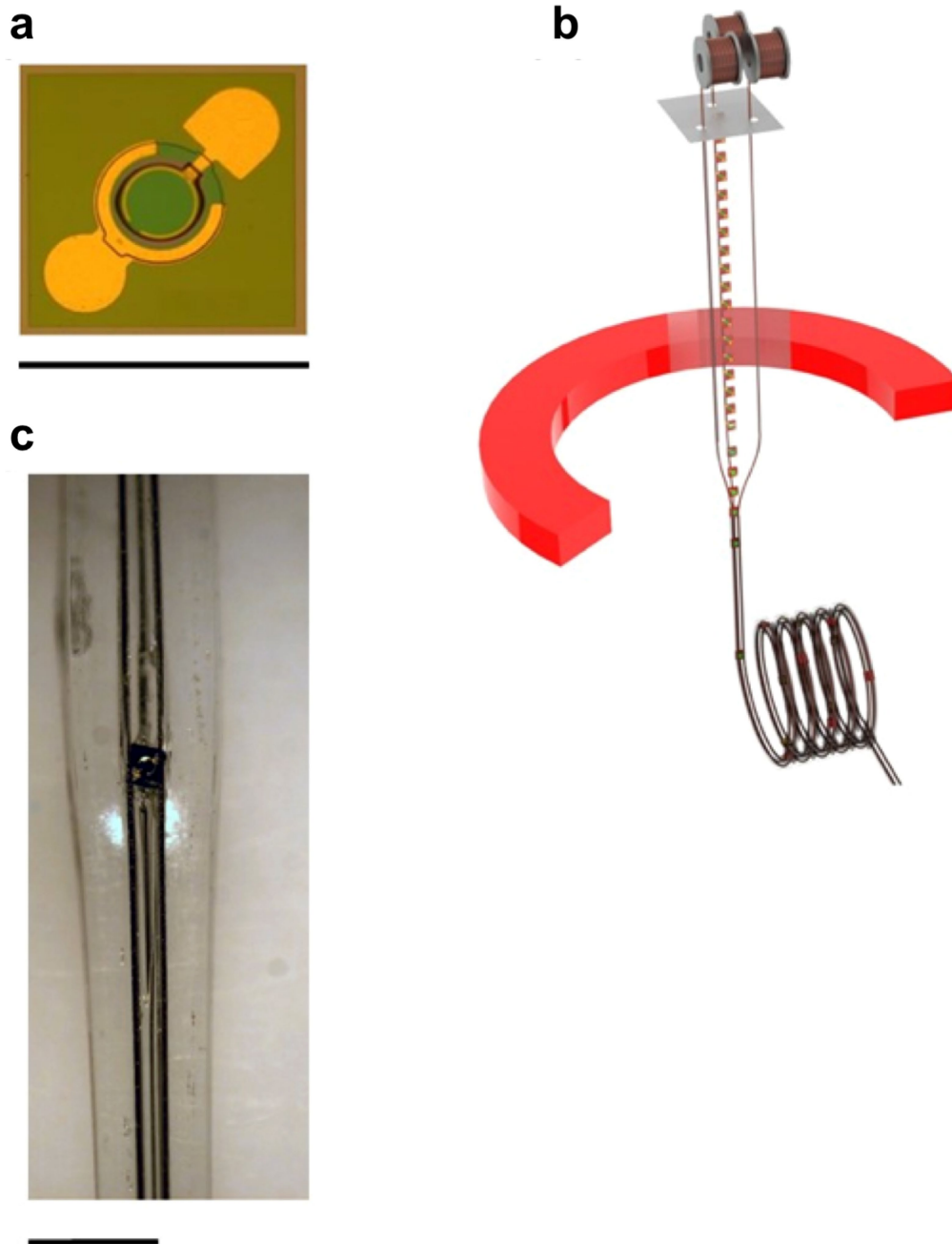
26. Hecht, E. *Optics* (Pearson Education, London, 2014).



Extended Data Fig. 1 | See next page for caption.

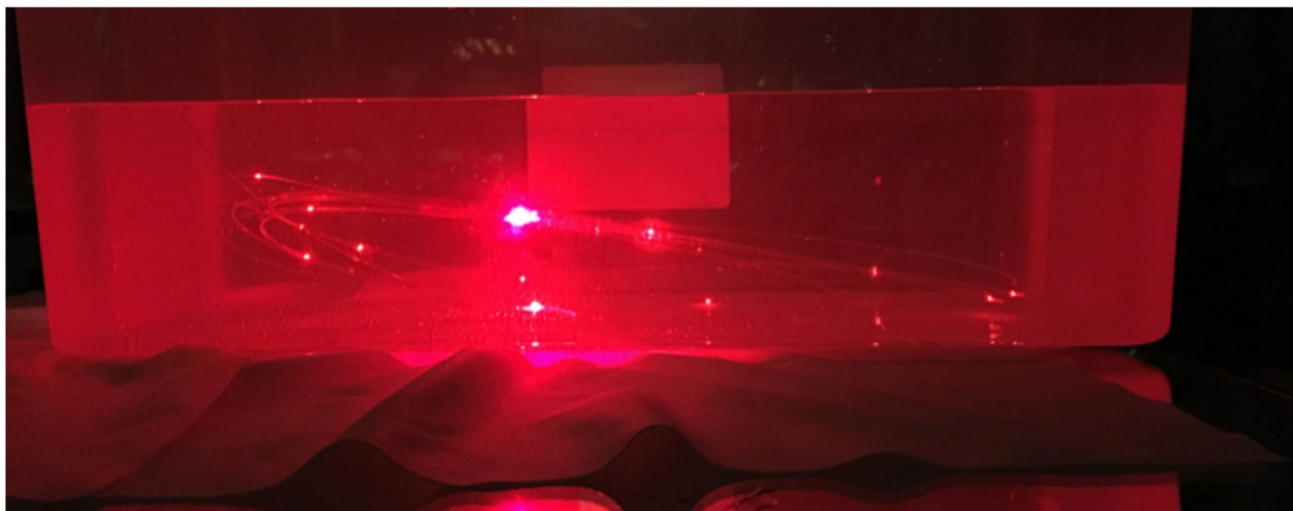
Extended Data Fig. 1 | Controlling device density in light-emitting fibres. **a**, Light-emitting fibres with low device density. (i) Photograph of light-emitting fibres containing blue-colour LEDs. The devices appear with a periodicity of $2,000 \pm 110$ mm. Scale bar, 50 cm. (ii) Illustration of the device density in the polymeric layer, which is placed in the middle of the preform, with a device separation of 1.25 mm. **b**, Higher density of devices in the fibre. (i) Photograph of light-emitting fibres containing blue-colour LEDs, where the devices appear every 370 ± 110 mm. This is the maximum linear density available with the given draw-down ratio (40) and device size for the single-layer (plane) architecture. Scale bar, 50 cm. (ii) Illustration of the maximal linear density of devices in the preform. The devices are placed side by side in a single plane. **c**, An alternative approach to increasing device density in fibres. (i) Illustration of the structure of the fibre cross-section, where light-emitting devices (blue shapes) are placed in two layers on top of each other, connected to metallic electrodes (red circles) for current delivery. The + and – signs represent the polarity of the wires when connected to the power supply.

(ii) Photograph of the resulting fibre, in which the devices appear every 173 ± 92 mm. Scale bar, 20 cm. (iii) Side view of the light-emitting fibre, showing the presence of three electrode wires. Scale bar, $600 \mu\text{m}$. **d**, Distance between devices in the fibre as a function of the distance between devices in the preform for a draw-down ratio of 40 and using LEDs. Solid lines show calculation results; black curve, single device layer in the preform; blue curve, two device layers in the preform; green curve, three device layers in the preform. Red circles represent measurements of inter-device spacing in the fibres. The dashed red line corresponds to the minimal distance between devices in the preform, which is equal to the size of the devices ($170 \mu\text{m}$). **e**, Distance between devices in the fibre as a function of the draw-down ratio. Solid curves show calculation results: black curve, single device layer with a spacing of $230 \mu\text{m}$ between devices in the preform; green curve, two device layers with a spacing of $230 \mu\text{m}$; blue curve, single device layer with a spacing of 1.25 mm. Red circles, measurements of inter-device spacing in the fibres. Error bars represent one standard deviation.

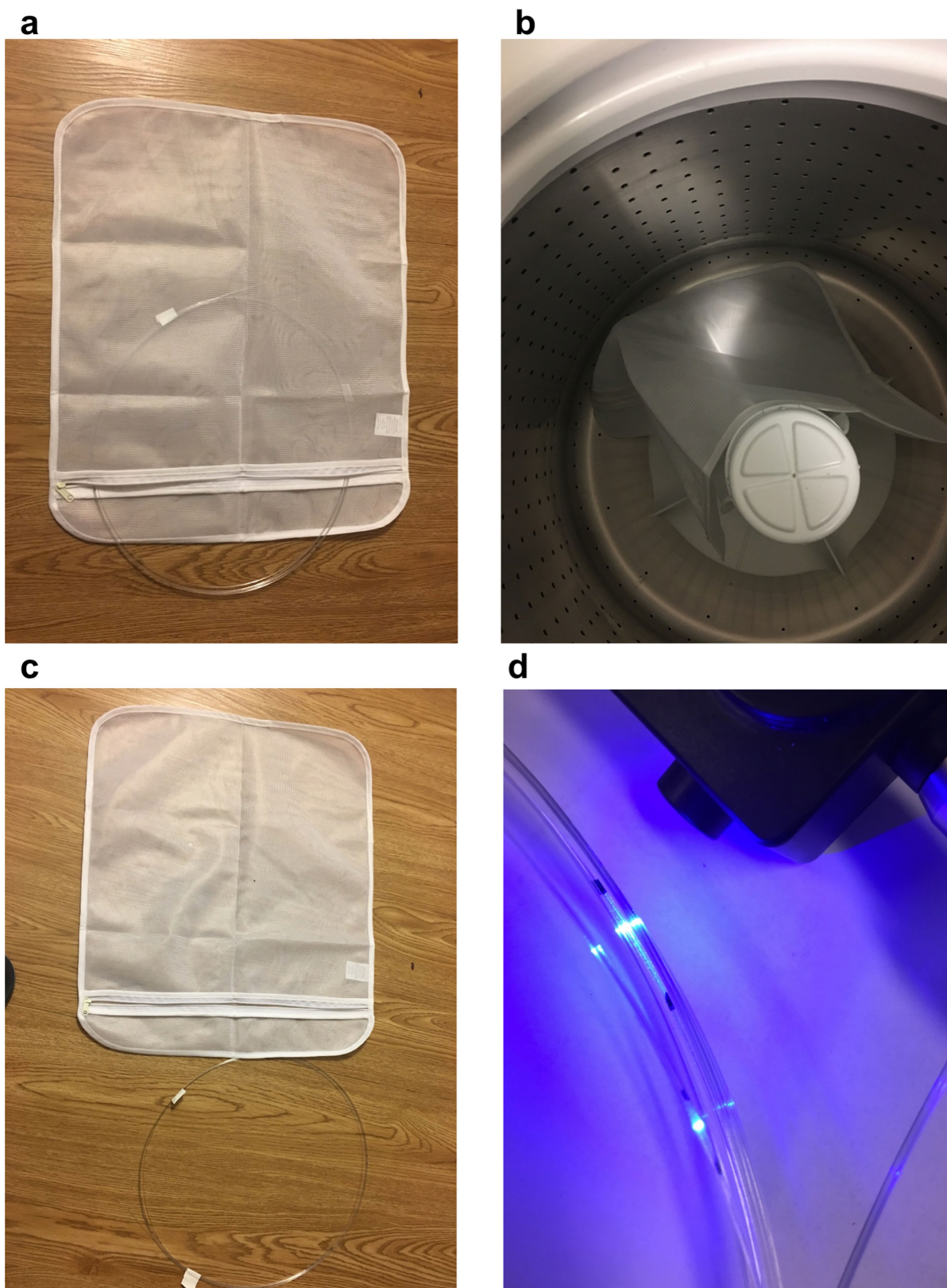


Extended Data Fig. 2 | Drawing of high-bandwidth photodetecting fibre. **a**, Optical micrograph of a commercial GaAs photodetecting device element. The central part is the device aperture, surrounded by two metallic contacts. Scale bar, $275\mu\text{m}$. **b**, Illustration of the preform drawing process for the photodetecting fibres. The contact to the devices

is established on the same side of the detectors, keeping the apertures of the devices uncovered by wires, whereas the third wire is placed behind the devices to prevent them from rotating during fibre drawing. **c**, Optical micrograph of the photodetecting fibres, showing a device embedded in the fibre. Scale bar, $600\mu\text{m}$.

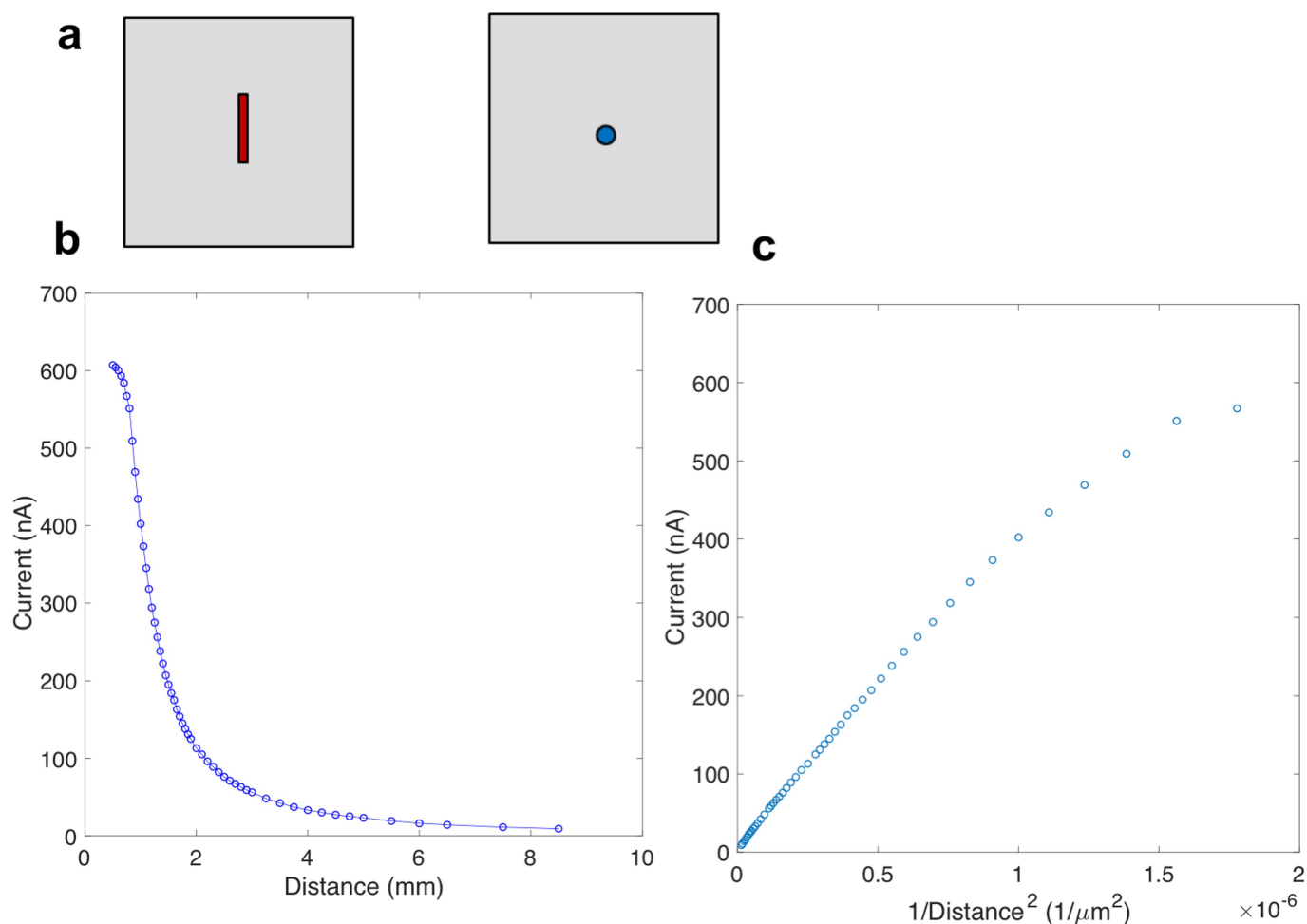


Extended Data Fig. 3 | Photograph of a light-emitting fibre immersed in a tank of water. The fibres are fully operational when immersed in water.



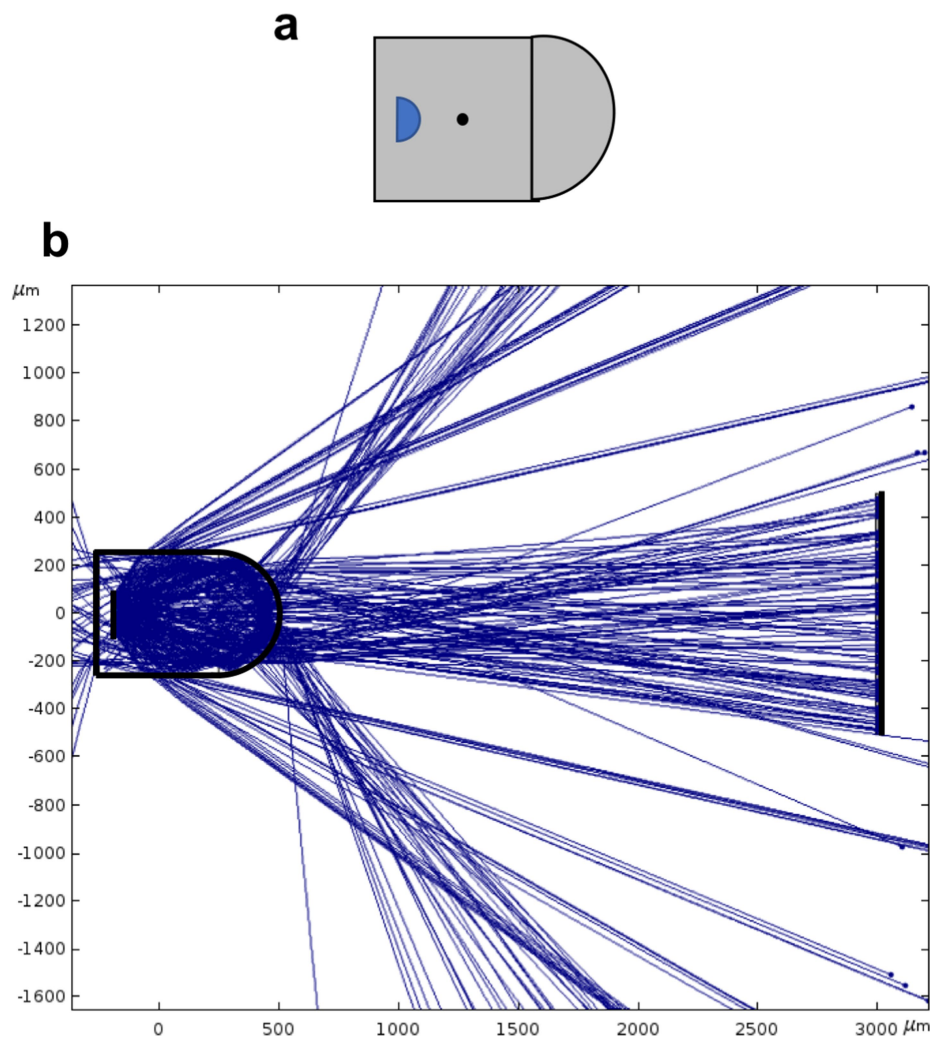
Extended Data Fig. 4 | Machine washing experiments with light-emitting fibres. **a**, A bunch of light-emitting fibres is placed in a water-permeable protective sack. **b**, The protective sack with the fibres is

placed in a household washing machine. **c**, Fibres and sack after a washing cycle. **d**, Fibre operation and light emission after the washing cycle.



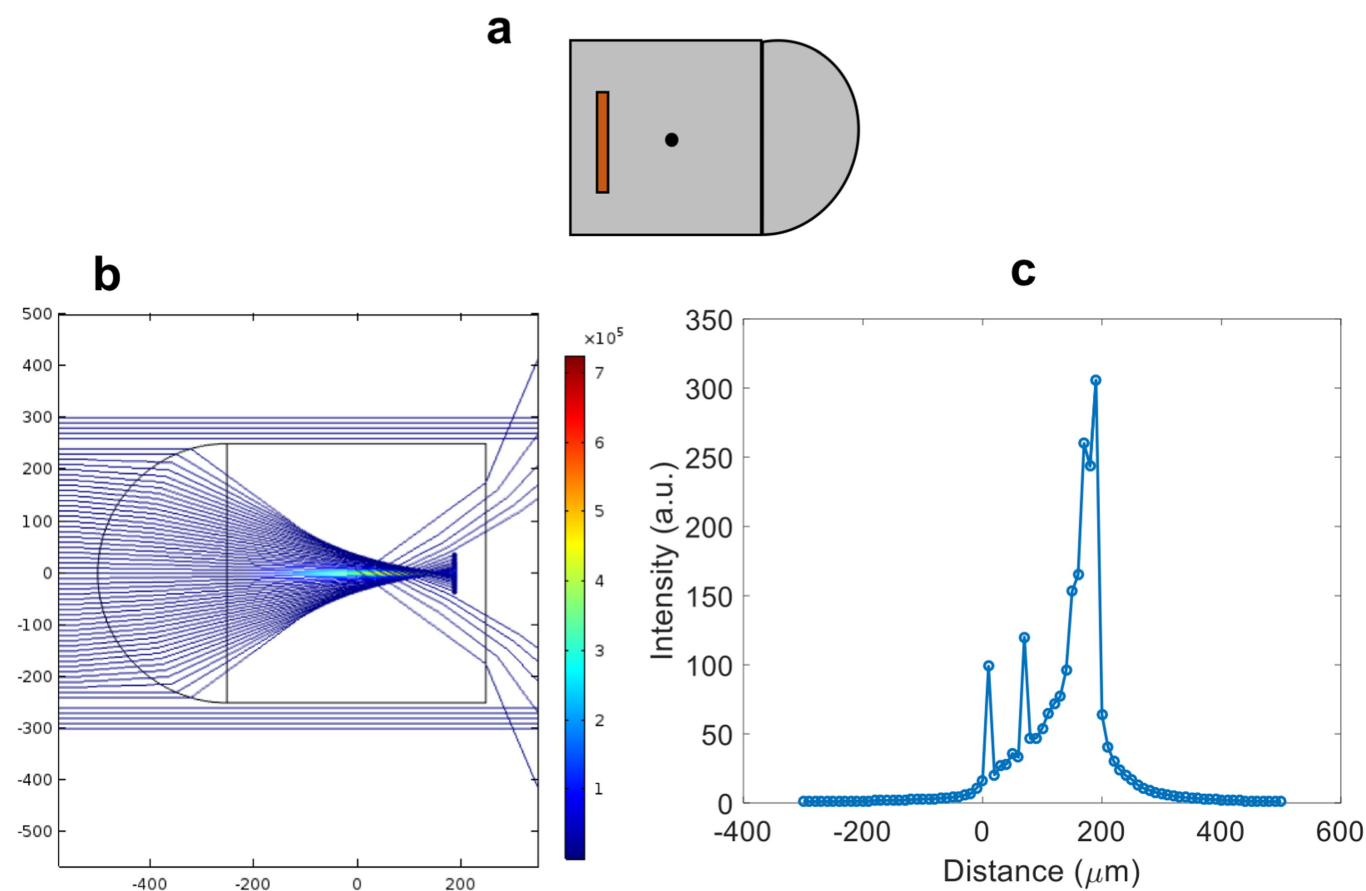
Extended Data Fig. 5 | Measurement of current registered by the photodetecting fibre as a function of the distance between the photoemitting and photodetecting fibres. a, Illustration of the experimental setup. Red rectangle, photodetector; blue circle, LED point source; grey square, PC cladding. **b,** Current registered by the photodetecting fibre versus its distance from the light-emitting fibre, obtained with the photodetecting fibre placed in front of a light-emitting fibre while varying the distance between them. **c,** Current versus the

inverse distance squared. The plot shows a linear dependence between the current and the inverse distance squared, which corresponds to the inverse-square law, at distances larger than 1 mm between the fibres. At shorter distances, deviation from the inverse-square law is observed. Several factors could contribute to this deviation, such as the finite sizes of the emitter and detector, the Lambertian profile of the emission and contact between the fibres at lower distances, which may have distorted the distance measurements between the fibres.



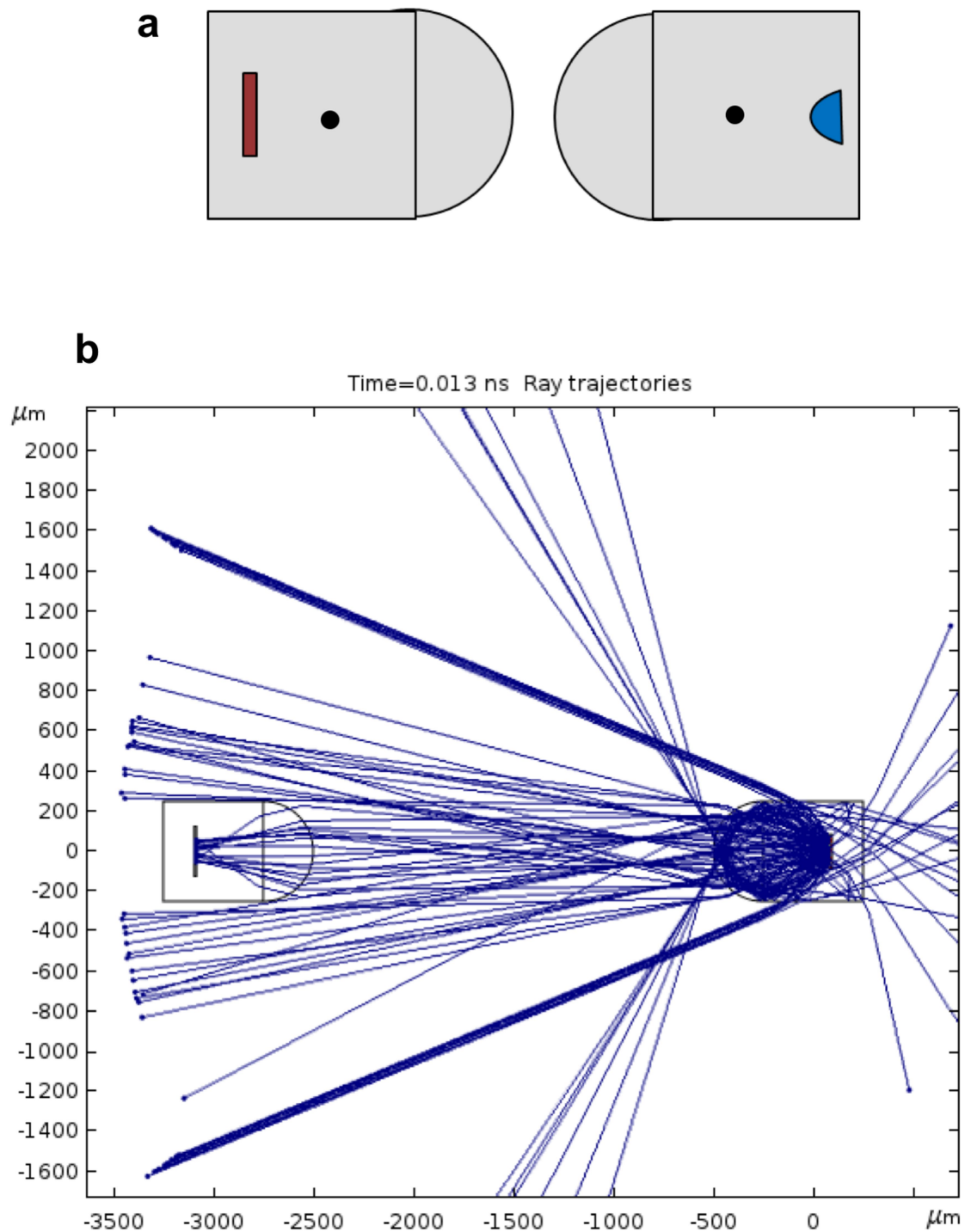
Extended Data Fig. 6 | Simulation of light-emitting fibre structure with a lens. a, Illustration of the fibre structure. Blue, LED; grey, PC cladding. Fibre size, $500\text{ }\mu\text{m} \times 500\text{ }\mu\text{m}$. The black dot shows the centre of the fibre. The radiation pattern was assumed to be Lambertian. **b,** Results of the ray-optics simulations, showing collimation of the light from the LED

when the device is placed at $190\text{ }\mu\text{m}$ from the centre of the fibre. The fibre structure is outlined by a black curve, and a general photodetector is plotted on the right-hand side of the figure. Blue lines represent optical rays emitted by the LED in the fibre.



Extended Data Fig. 7 | Simulation of photodetecting fibre structure with a lens. **a**, Illustration of the fibre structure. Red, photodetector; grey, PC cladding. Fibre size, $500\mu\text{m} \times 500\mu\text{m}$. The centre of the fibre is denoted by a black dot. **b**, Results of the ray-optics simulations, showing focusing of the collimated external light on the photodetecting device.

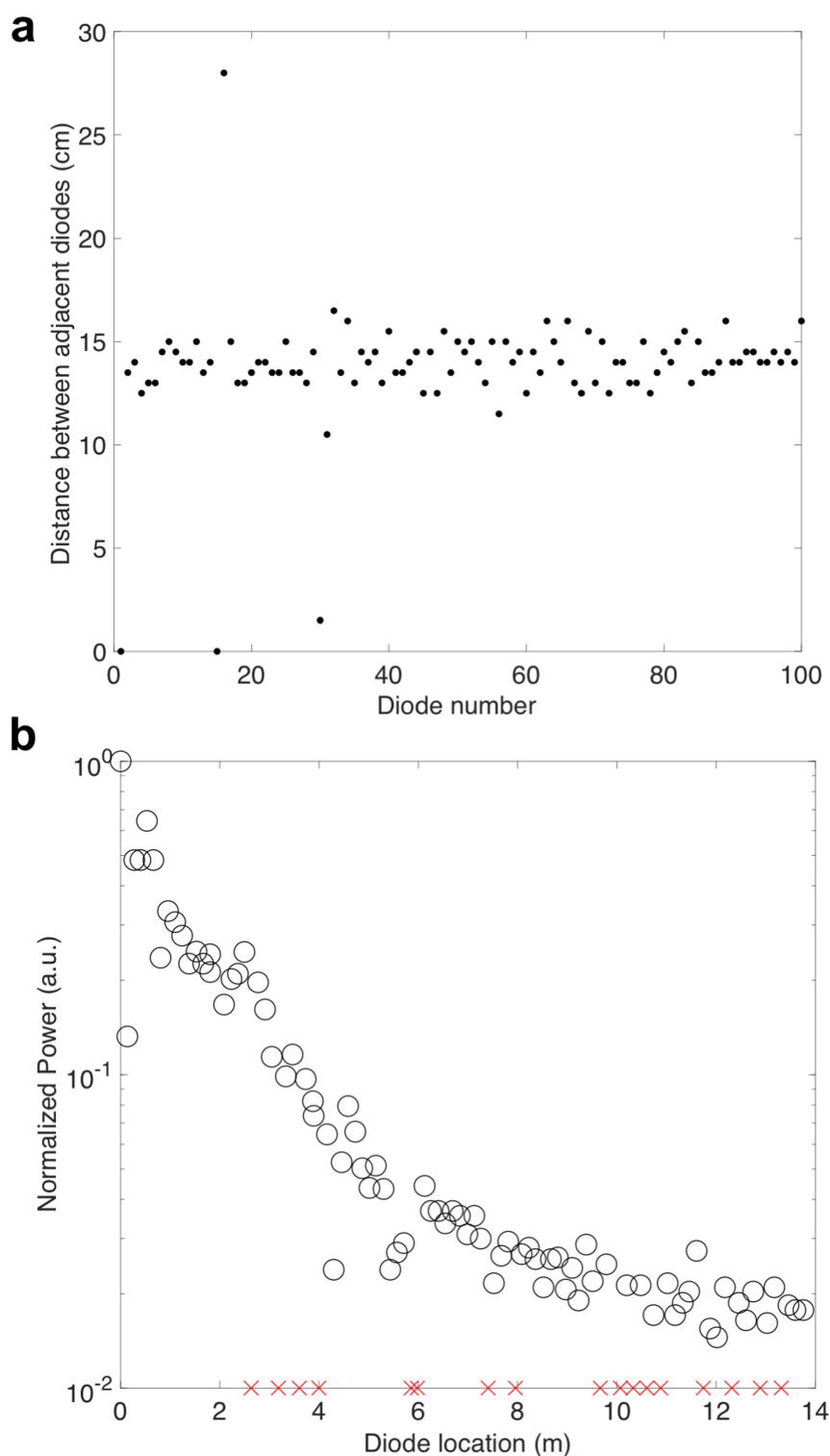
The results were obtained with the device placed $190\mu\text{m}$ away from the centre of the fibre. Axis units, μm . **c**, The intensity of the illumination as a function of the location of the device in the fibre. The maximal intensity is achieved at $190\mu\text{m}$ from the centre of the fibre.



Extended Data Fig. 8 | Simulation of a lensed communication system containing a light-emitting fibre and a photodetecting fibre.

a, Illustration of the fibre system structure. Red, photodetector; blue, light-emitting device; grey, PC cladding. Fibre size, $500\text{ }\mu\text{m} \times 500\text{ }\mu\text{m}$. The black

dot denotes the centre of the fibre. **b**, Results of the ray-optics simulations that show collimation of the emitted light and focusing of the light on the photodetecting device, with the devices placed $190\text{ }\mu\text{m}$ from the centre of the fibre.



Extended Data Fig. 9 | Characterization of a typical preform draw.
a, Measured distance between adjacent LEDs in the drawn fibres. The diodes were arranged in two parallel arrays in the preform, which was drawn with a draw-down ratio of 33. **b**, Optical power characterization of the LEDs in the drawn fibre. The power was normalized with the power of

the brightest diode, which was located adjacent to the power source. Non-operational LEDs are marked by a red cross. The emitted power decays as the voltage drops on the wires in the fibre for devices located away from the power source.

Extensive loss of past permafrost carbon but a net accumulation into present-day soils

Amelie Lindgren^{1,2*}, Gustaf Hugelius^{1,2} & Peter Kuhry^{1,2}

Atmospheric concentrations of carbon dioxide increased between the Last Glacial Maximum (LGM, around 21,000 years ago) and the preindustrial era¹. It is thought that the evolution of this atmospheric carbon dioxide (and that of atmospheric methane) during the glacial-to-interglacial transition was influenced by organic carbon that was stored in permafrost during the LGM and then underwent decomposition and release following thaw^{2,3}. It has also been suggested that the rather erratic atmospheric $\delta^{13}\text{C}$ and $\Delta^{14}\text{C}$ signals seen during deglaciation^{1,4} could partly be explained by the presence of a large terrestrial inert LGM carbon stock, despite the biosphere being less productive (and therefore storing less carbon)^{5,6}. Here we present an empirically derived estimate of the carbon stored in permafrost during the LGM by reconstructing the extent and carbon content of LGM biomes, peatland regions and deep sedimentary deposits. We find that the total estimated soil carbon stock for the LGM northern permafrost region is smaller than the estimated present-day storage (in both permafrost and non-permafrost soils) for the same region. A substantial decrease in the permafrost area from the LGM to the present day has been accompanied by a roughly 400-petagram increase in the total soil carbon stock. This increase in soil carbon suggests that permafrost carbon has made no net contribution to the atmospheric carbon pool since the LGM. However, our results also indicate potential postglacial reductions in the portion of the carbon stock that is trapped in permafrost, of around 1,000 petagrams, supporting earlier studies⁷. We further find that carbon has shifted from being primarily stored in permafrost mineral soils and loess deposits during the LGM, to being roughly equally divided between peatlands, mineral soils and permafrost loess deposits today.

It has been proposed previously that the global terrestrial carbon stock increased from the LGM to the present day^{8,9}. However, these studies did not explicitly consider permafrost or deep-soil carbon stocks, which would have caused them to underestimate soil carbon storage in certain regions during the LGM. Moreover, these studies did not look at the potential loss of permafrost-trapped soil carbon during deglaciation—information that is needed to resolve atmospheric isotope signals². Modern permafrost soils store considerable amounts of organic carbon¹⁰; because of this, it has been suggested that the larger area of permafrost during the LGM¹¹ and the greater extent of permafrost loess deposits⁶ led to higher-than-present soil carbon storage at that time. In the absence of empirical reconstructions of carbon storage within the LGM permafrost zone, estimates have relied on model outputs and endmember calculations^{5,12}. However, present Earth system models (ESMs) cannot represent the key processes of glacial-to-interglacial CO_2 dynamics owing to uncertain parameterization of peat and permafrost carbon dynamics¹³. Although ESMs are improving rapidly^{12,14} and hold the potential of projecting forwards in time, they must still rely on empirical palaeontological data for validation of past glacial cycles.

Here we combine an extensive range of empirical data on past environments to explore and categorize the LGM permafrost landscape,

and to compare it with the present-day landscape in the same region. We define carbon stored in permafrost itself as inert, and compare how this inert fraction of the total carbon stock changed from the LGM to the present.

As the basis of our calculations, we adapted LGM biome reconstructions^{15–17} to delineate areas that were dominated by tundra, forest and steppe biomes, all of which encompassed a variety of plant communities (Fig. 1a and Extended Data Table 1). Within these broader categories, we differentiated lowland and alpine zones¹⁸, as well as zones with lower or higher peatland coverage (Fig. 1b) according to findings of buried peat and counts of *Sphagnum* moss spores in pollen assemblages (Extended Data Fig. 1). To reconstruct typical carbon stocks for these past regions, we compared them with modern-day tundra, taiga and steppe within the present permafrost zone¹⁹. By assuming a comparable magnitude and variability of landscape carbon stocks between past and present biomes, we estimated LGM carbon stocks down to a depth of 3 m on the basis of present-day data from North America²⁰ for taiga and tundra, and from the Tibetan plateau²¹ for steppe (Extended Data Table 2). Alpine regions with steep mountain slopes were reconstructed separately. We calculated a mineral-soil carbon stock of 790 Pg for the whole LGM permafrost region, mainly from carbon-rich tundra soils (Table 1). A striking difference between past and modern permafrost environments is the apparent lack of peatlands at LGM times. Extensive databases and previous research notwithstanding, records of northern peatlands older than 16.5 thousand years are scarce²², indicating limited peatland development during LGM times. Consequently, we reconstructed an LGM peatland carbon stock of only 30 Pg.

Because sea levels during the LGM were lower than today, the LGM landscape included areas of exposed sea shelves. We included 0–3 m depth of carbon stocks from these areas in our overall biome reconstructions, amounting to an additional carbon storage of 220 Pg. We assume that these shelves, which have since been inundated by the sea, have retained the carbon accumulated during glacial times. Very limited data are available for sea-shelf carbon stocks²³, but we assume that any carbon that may have been lost through sub-sea permafrost degradation and microbial decomposition has been compensated by fresh sediment deposition. Another important landscape element during glacial times was the ice sheets themselves. Preglacial landscapes might have been partially preserved under cold-based sheets²⁴ (Extended Data Fig. 2), and we reconstructed an inert LGM carbon stock of 120 Pg from these subglacial areas. We assume no changes to carbon stocks beneath the still-existent Greenland Ice Sheet (50 Pg).

Extensive areas with loess sequences in the Northern Hemisphere formed over several glacial periods, including the LGM, and it has been proposed that their accumulative genesis resulted in carbon-rich deposits across the past permafrost zone⁶, similar to the Beringian Yedoma deposits (Fig. 1b; Yedoma deposits are organic- and ice-rich permafrost of Pleistocene age). The depth and carbon stocks of these deposits are included in our LGM carbon estimate. However, we conclude that deposits pre-dating the coldest interval of the last glacial period (marine isotope stages 4–2)—which lie outside the present northern permafrost

¹Department of Physical Geography, Stockholm University, Stockholm, Sweden. ²Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden. *e-mail: amelie.lindgren@natgeo.su.se

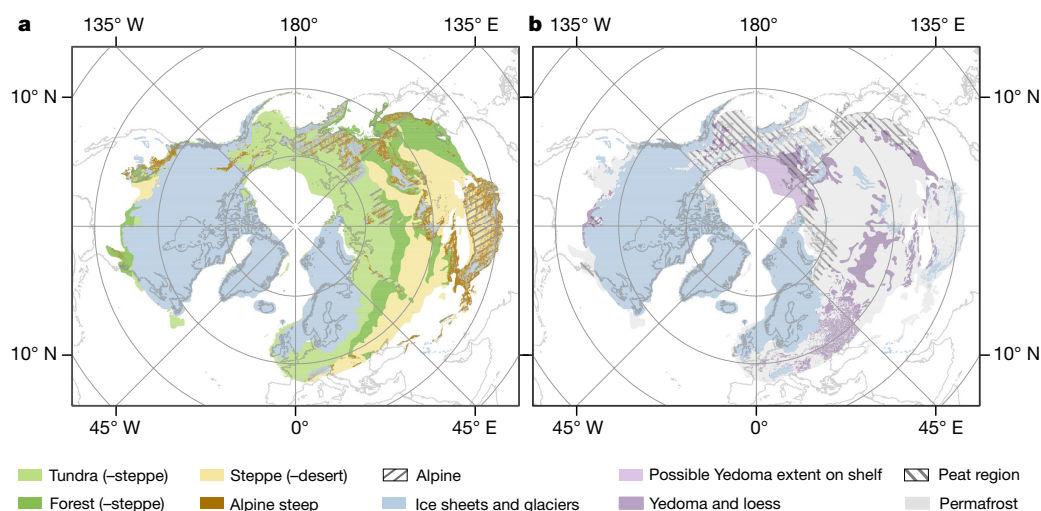


Fig. 1 | Reconstructed LGM environment. **a**, ‘Mega’ biomes (defined in Extended Data Table 1) and ice-sheet extents in the LGM northern permafrost region. In parentheses are assumed secondary occurrences of other biome types within broader mega biomes. **b**, Spatial extents of permafrost, Yedoma, loess and peat regions in the LGM northern permafrost region. The Yedoma and loess deposits include the majority

region—were affected by (repeated) thaw in warm interglacial and interstadial periods before the LGM. This resulted in a substantial depletion of their initial high carbon stocks before the LGM, and we reconstructed an additional storage of 366 Pg C (range 56–725 Pg) during the LGM, which is far less than the 1,000 Pg suggested previously⁶. We assumed that other deep permafrost carbon stocks—such as those on the Siberian shelf, in deltas, and in the current Yedoma region (Table 1)—were constant between the LGM and the present, with small changes in the inert component. We have not explicitly considered stocks in other deep Quaternary deposits⁸, but assume that they have remained constant.

Surprisingly, we find that the total estimated soil carbon stock for the LGM northern permafrost region is smaller than the estimated present-day storage (in both permafrost and non-permafrost soils), if the same areas are compared (2,300 Pg and 2,700 Pg, respectively; Table 1 and Supplementary Table 4). We assessed uncertainties in our reconstructions and determined a plausible maximum and minimum range of LGM permafrost carbon stocks of between 1,680 Pg and 2,860 Pg (Table 1; present-day range 2,440–3,070 Pg). We used a range

of our deep deposits, with additional carbon storage occurring in deltas (extent not reconstructed). The peat region depicts areas in which we reconstruct a higher LGM peat coverage of 5%, compared with 1% in other regions (that is, areas within the LGM permafrost region but outside of the peat region).

of scenarios for uncertainty quantification because the nature of this LGM reconstruction precludes traditional statistical quantification of variance or uncertainty. These error ranges represent the main reconstruction uncertainties—that is, the average carbon density of different LGM ecosystems, the distribution of biomes, the areal coverage of peatlands, storage in deep loess deposits and the possible storage of carbon beneath ice sheets. We provide a longer discussion about uncertainties in Supplementary Information.

The net gain of carbon from 2,300 Pg in the LGM to 2,700 Pg at present does not imply gradual carbon accumulation following post-glacial warming and permafrost thaw. Instead, from LGM times to the present there is evidence for a geographic shift in carbon storage. There has been a net transfer of carbon from mineral permafrost soils and subglacial and deep deposits, via the atmosphere, into thawed mineral soil and both frozen and thawed organic soils (Fig. 2). Previous empirical studies have also noted an increase in global terrestrial carbon storage, including in vegetation, over the same period⁸. Carbon storage in both vegetation and soils may also have been higher than at present at some point during the Holocene epoch^{8,9}. Our estimate

Table 1 | Estimated carbon pools (in Pg C) for the LGM and present day

	LGM	Range	LGM inert	Range	Present	Range	Present inert	Range
Mineral soil (0–3 m)	790	269–1143	574	177–838	1,084	840–1,366		
(of which permafrost region)					(589)	367–811	439	270–608
Peatland (0–3 m)	30	16–180	20	11–121	550	457–683		
(of which permafrost region)					(153)	91–215	127	75–179
Shelf (0–3 m)	220	64–252	164	41–183	220	64–251		
(of which permafrost region)					(122)	33–130		
Deep deposits: Yedoma (>3 m)	741	610–884	741	610–884	741	624–869		
(of which permafrost region)					(718)	601–846	669	564–785
Deep deposits: loess (>3 m)	366	56–725	366	56–725	48	9–92		
Deltas	91	37–135	91	37–135	91	37–135	69	31–107
Large lakes	2				45			
Subglacial	170	117–225	171	117–225	(48)		48	
Total	2,319	1,677–2,867	2,035	1,456–2,522	2,736	2,436–3,073	1,283	1,077–1,495

This table summarizes and compares carbon storage during the LGM and present-day carbon storage within the same region, including the permafrost-inert part of each stock. Woody litter is not included. Deep deposits are separated into Yedoma and loess, where the former includes Yedoma carbon storage on sea shelves. Plausible range scenarios for the upper three metres of soil are estimated for the LGM (see Methods). The range for the deep deposits includes both well constrained error estimates from present deposits and a statistical analysis of the depth distribution for additional LGM permafrost loess deposits. The total ranges were calculated by additive error propagation. For more details and descriptions of how the present-day carbon pool was quantified, see Methods, Supplementary Information and Supplementary Table 4.

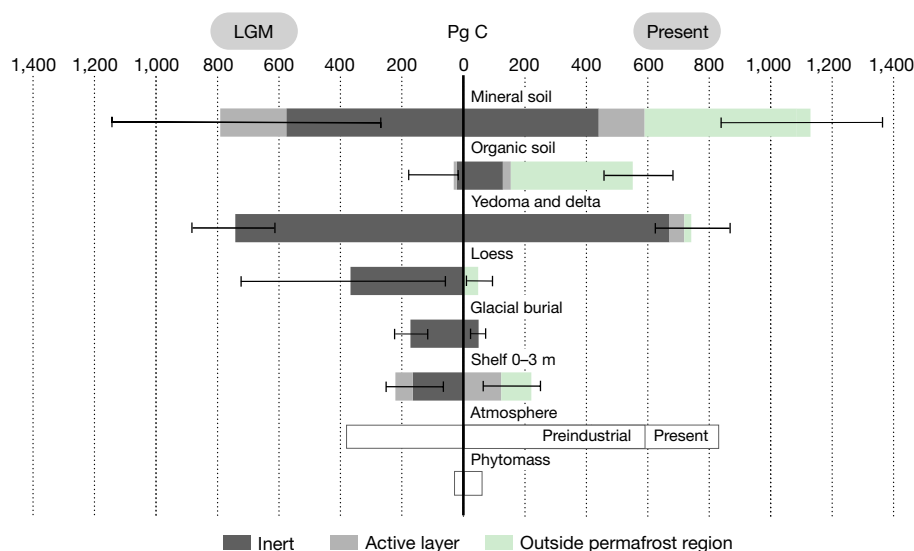


Fig. 2 | Sizes of carbon stocks in the LGM northern permafrost region during the LGM and at present. Separate stocks are reported for total carbon (shown by the total length of each bar), separated into inert carbon (dark grey; defined as permanently frozen carbon), active-layer carbon (light grey; defined as carbon within the seasonally thawed layer of soil above the permafrost table), and unfrozen carbon outside the modern permafrost region (light green). Carbon stocks in the atmosphere and in

plant material (phytomass) from the LGM northern permafrost region are included in the figure for reference, but are not included in the total carbon stocks discussed in the text. The carbon stocks of large lakes are not shown separately, but are included in the mineral-soil carbon as unfrozen (not visible for the LGM). Error bars are ranges around the total carbon stocks, specified in Table 1 and Methods.

of the LGM inert permafrost carbon stock (2,000 Pg C) is somewhat lower than suggested previously^{5,12} (2,300 Pg), but is much larger than the present inert permafrost carbon stock (1,300 Pg). This agrees with previous findings of postglacial permafrost carbon remobilization³. Known thermokarst events (ground subsidence caused by the melting of massive ground ice) in the current Yedoma region postdate the LGM³, and our review of loess sequences located outside the present-day permafrost region shows no evidence of thermokarst at LGM times. The thaw of permafrost deposits following postglacial warming would have exposed organic matter to decomposition, resulting in the release of carbon depleted relative to the atmosphere in both $\Delta^{14}\text{C}$ (on account of its greater age) and $\delta^{13}\text{C}$ (caused by the preference of light-carbon uptake by plants and further fractionation during decomposition). These isotopic properties of the thawed material fit well with the development of atmospheric isotopic signals preserved in ice cores¹⁴. Decomposition of a putative, but highly uncertain, old and inert subglacial carbon stock following the gradual retreat of the large Northern Hemisphere ice sheets could also have contributed to these observed changes in atmospheric isotope composition.

While widespread thermokarst formation occurred in the Yedoma region during the Late Glacial and Early Holocene³, new land areas became available for soil development following deglaciation of the Laurentide and Fennoscandian ice sheets²⁵. This also corresponds to a time period of widespread peatland formation in the Northern Hemisphere²². The postglacial environment has changed dramatically, and parts of the landscape that were previously occupied by the relatively dry tundra (or steppe–tundra) are today covered by peat soils²⁶. These organic soils represent a considerable portion of the present Northern Hemisphere carbon stocks²⁷. Therefore, the glacial–interglacial transition seems to correspond to a period involving a depletion of permafrost carbon stocks, while at the same time new stocks started to accumulate in other soils²⁵. In the present discontinuous permafrost region, the aggradation of new permafrost into previously accumulated peat deposits following Late Holocene cooling has resulted in the formation of new inert carbon storage. Considering all of these lines of evidence, it is possible that around 1,000 Pg of inert carbon became activated during the deglaciation.

By their nature, reconstructions of past ecosystems and environments rely on assumptions that are highly uncertain and difficult to

validate. Through the use of analogous modern-day landscapes, we assume a comparable magnitude and variability of soil carbon stocks between past and present. However, LGM plant communities existed in forms we do not see today, such as the steppe–tundra biome with its co-dominance of steppe and tundra plant species²⁸. Plant productivity, and carbon input to soils, was probably lower owing to lower atmospheric CO_2 concentrations. This may have been especially important in forest systems, as low CO_2 levels favoured more open vegetation⁹, but the response of cold-region ecosystems to variable ambient CO_2 levels remains uncertain even today, and experiments show that changes in plant productivity under different CO_2 concentrations do not necessarily change ecosystem carbon storage²⁹. Some authors have suggested that fast biochemical cycling because of the presence of large grazers (Extended Data Fig. 3), and elevated dust loads supplying fresh nutrients, enabled productive, and carbon-rich, ecosystems during the LGM^{6,30}. This idea is supported by the high carbon stocks observed in the preserved LGM Yedoma region, similar to those we reconstruct, but it is unclear to what extent the preserved Yedoma is representative of the vast LGM region.

The extent of peatlands and wetlands during the LGM is also highly uncertain. Our reconstructions are based partly on the occurrence of *Sphagnum* spores, but this could have led us to miss minerotrophic fens characterized by graminoids and brown mosses. The scenario-based analysis also shows a skewed error range towards a possibly higher carbon storage. On the other hand, few deep peat deposits are dated to LGM times. Most of the stratigraphic evidence points towards thin (less than 40 cm) peat layers, which in our reconstructions are included in upland mineral-soil reconstructions. Speculation regarding a possible widespread oxidation of LGM peat deposits before the onset of post-glacial peatland development lies outside the scope of this empirically based study.

To alleviate some of these uncertainties and to further refine estimates of glacial to interglacial carbon-stock dynamics, further research is needed. Specifically, we propose further research into potential subglacial carbon storage, the initial stock and fate of carbon on inundated sea shelves, the potential extent of peatlands and wetlands during the LGM, and the effect of both lower and higher atmospheric CO_2 levels on tundra and boreal forest productivity and carbon turnover.

Our reconstructions suggest that the loss of more than 10 million square kilometres of northern permafrost area since the LGM has resulted in the net addition of several hundred petagrams of carbon into present-day soils. Nevertheless, postglacial warming and permafrost thaw resulted in an initial large loss of inert carbon, which may have approached 1,000 Pg. This initial loss of carbon was compensated by carbon accumulation in permafrost-free mineral soils, in deglaciated terrain, and in peatlands. More research is needed to disentangle transient changes during the early stages of the last deglaciation and postglacial warming. We stress that the response of the LGM permafrost carbon stock to thaw may not be a good analogue for the fate of the present permafrost stock, which has a different composition to that of the past.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0371-0>.

Received: 30 August 2017; Accepted: 14 May 2018;

Published online 1 August 2018.

- Schmitt, J. et al. Carbon isotope constraints on the deglacial CO₂ rise from ice cores. *Science* **336**, 711–714 (2012).
- Crichton, K. A., Bouttes, N., Roche, D. M., Chappellaz, J. & Krinner, G. Permafrost carbon as a missing link to explain CO₂ changes during the last deglaciation. *Nat. Geosci.* **9**, 683–686 (2016); corrigendum **9**, 795 (2016).
- Walter, K. M., Edwards, M. E., Grosse, G., Zimov, S. A. & Chapin, F. S. III. Thermokarst lakes as a source of atmospheric CH₄ during the last deglaciation. *Science* **318**, 633–636 (2007).
- Broecker, W. & Barker, S. A. 190‰ drop in atmosphere's $\delta^{14}\text{C}$ during the 'Mystery Interval' (17.5 to 14.5 kyr). *Earth Planet. Sci. Lett.* **256**, 90–99 (2007).
- Ciais, P. et al. Large inert carbon pool in the terrestrial biosphere during the Last Glacial Maximum. *Nat. Geosci.* **5**, 74–79 (2011).
- Zimov, N. S. et al. Carbon storage in permafrost and soils of the mammoth tundra-steppe biome: role in the global carbon budget. *Geophys. Res. Lett.* **36**, L02502 (2009).
- Anthony, K. M. W. et al. A shift of thermokarst lakes from carbon sources to sinks during the Holocene epoch. *Nature* **511**, 452–456 (2014).
- Adams, J. M., Faure, H., Faure-Denard, L., McGlade, J. M. & Woodward, F. I. Increases in terrestrial carbon storage from the Last Glacial Maximum to the present. *Nature* **348**, 711–714 (1990).
- Prentice, I. C., Harrison, S. P. & Bartlein, P. J. Global vegetation and terrestrial carbon cycle changes after the last ice age. *New Phytol.* **189**, 988–998 (2011).
- Hugelius, G. et al. Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **11**, 6573–6593 (2014).
- Lindgren, A., Hugelius, G., Kuhry, P., Christensen, T. R. & Vandenbergh, J. GIS-based maps and area estimates of Northern Hemisphere permafrost extent during the Last Glacial Maximum. *Permafrost Periglacial Process.* **27**, 6–16 (2016).
- Zhu, D. et al. Simulating soil organic carbon in yedoma deposits during the Last Glacial Maximum in a land surface model. *Geophys. Res. Lett.* **43**, 5133–5142 (2016).
- Brovkin, V. et al. Comparative carbon cycle dynamics of the present and last interglacial. *Quat. Sci. Rev.* **137**, 15–32 (2016).
- Kleinen, T., Brovkin, V. & Munhoven, G. Modelled interglacial carbon cycle dynamics during the Holocene, the Eemian and Marine Isotope Stage (MIS) 11. *Clim. Past* **12**, 2145–2160 (2016).
- Grichuk, V. P. in *Late Quaternary Environments of the Soviet Union* (ed. Velichko, A. A.) 155–178 (Longman, London, 1984).
- Dyke, A. S. Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. *Geogr. Phys. Quat.* **59**, 211–262 (2005).
- Baryshnikov, G. A. & Markova, A. K. in *Paleoclimates and Paleoenvironments of Extra-tropical Regions of the Northern Hemisphere. Late Pleistocene–Holocene. Atlas-monograph* (ed. Velichko, A. A.) 79–87 (GEOS, Moscow, 2009).
- Gruber, S. Derivation and analysis of a high-resolution estimate of global permafrost zonation. *Cryosphere* **6**, 221–233 (2012).
- Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
- Hugelius, G. et al. The northern circumpolar soil carbon database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5**, 3–13 (2013).
- Ding, J. et al. The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. *Glob. Change Biol.* **22**, 2688–2701 (2016).
- Macdonald, G. M. et al. Rapid early development of circumarctic peatlands and atmospheric CH₄ and CO₂ variations. *Science* **314**, 285–288 (2006).
- Schuur, E. A. G. et al. Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015).
- Kleman, J. & Hattestrand, C. Frozen-bed Fennoscandian and Laurentide ice sheets during the Last Glacial Maximum. *Nature* **402**, 63–66 (1999).
- Harden, J. W., Sundquist, E. T., Stallard, R. F. & Mark, R. K. Dynamics of soil carbon during deglaciation of the Laurentide Ice Sheet. *Science* **258**, 1921–1924 (1992).
- Loisel, J. et al. Insights and issues with estimating northern peatland carbon stocks and fluxes since the Last Glacial Maximum. *Earth Sci. Rev.* **165**, 59–80 (2017).
- Yu, Z. C. Northern peatland carbon stocks and dynamics: a review. *Biogeosciences* **9**, 4071–4085 (2012).
- Willerslev, E. et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).
- Norby, R. J. & Zak, D. R. Ecological lessons from free-air CO₂ enrichment (FACE) experiments. *Annu. Rev. Ecol. Evol. Syst.* **42**, 181–203 (2011).
- Yurtsev, B. A. The Pleistocene 'tundra-steppe' and the productivity paradox: the landscape approach. *Quat. Sci. Rev.* **20**, 165–174 (2001).

Acknowledgements This study was funded by a French–Swedish cooperation grant from the Swedish Research Council (349-2012-6293), and by the European Union (EU) Joint Programming Initiative (JPI) Climate Constraining Uncertainties in the Permafrost–Climate Feedback (COUP) consortium. G.H. acknowledges a Marie Curie Skłodowska and Swedish Research Council International Career Grant (INCA; no. 330-2014-6417). We also acknowledge J.-O. Persson at the Department of Mathematics, Stockholm University, who assisted with statistical consulting, and the Bolin Centre for Climate Research for hosting the data.

Reviewer information Nature thanks V. Brovkin, N. Pastick and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.L. was responsible for data collection, digitization of data, and analysis of data from past environments, excluding deep loess deposits. P.K. was responsible for the review and analysis of deep loess deposits. G.H. was responsible for estimates of modern carbon pools. All authors contributed substantially to formulating the research idea, interpreting the results and writing the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0371-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0371-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Biome reconstructions. Empirical map reconstructions of LGM biomes—based on pollen, plant macrofossils and/or faunal remains^{15–17}—were reviewed, digitized and compared to produce an aggregate biome-reconstruction map for the LGM permafrost region. The individual classes were harmonized to a common and simplified biome classification scheme (Extended Data Table 1). This harmonization required us to generalize biomes into the following broader categories: tundra (–steppe), forest (–steppe), and steppe (–desert), where each category in parenthesis defines secondary vegetation types. Findings of LGM megafaunal remains^{17,31,32} were briefly reviewed as a complement to the biome reconstructions (see Supplementary Information and Extended Data Fig. 3). The resulting map was compared with independent data points of biomized pollen counts^{9,33–36} (Kappa 0.85; see Supplementary Information and Extended Data Fig. 4).

Following the same overall procedure, we harmonized reconstructions of various alpine environments into alpine mega biomes (Extended Data Table 1). Moreover, using present-day topographic data (we assume no major changes in topography since the LGM), we categorized additional areas as alpine if they display a ruggedness index equal to or larger than 4. For more information about the procedure and data used for this classification, see ref. 18. By categorizing areas as rugged, a reconstructed tundra (–steppe) area becomes alpine tundra (–steppe), and so on. This scheme also allows mountain ranges such as the Alps to be identified as alpine.

Steep areas in cold climates are characterized by thin soils, talus formations and limited vegetation coverage, while valley floors may accumulate more carbon-rich soils (see, for example, ref. 37). On the basis of terrain slope³⁸ for the (mixed) alpine and topographically rugged areas, we separated steep areas from valley floors with a slope threshold of 4 degrees³⁹ (see Supplementary Information for details).

Peatlands. We digitized a reconstructed possible LGM peatland region on the basis of a range of evidence indicative of the presence of peatland (Extended Data Fig. 1). This included previous reconstructions of LGM peatlands^{40,41}, local to regional studies of peat^{42,43} and peaty (with O-horizons between 10 cm and 40 cm thick) deposits^{44–49}, and palynological data of *Sphagnum* spores^{34,50–64}. The delineation of the possible peatland regions was done by hand, including previously reported regions, and generally accepting *Sphagnum* spore counts greater than 1% (ref. 40) with indicative age control. Spore percentages below 1% were not accounted for unless they occurred in relatively close proximity. We do not take ruggedness into account when estimating the extent of this region. Further methods are available in Supplementary Information.

To estimate peatland extent within the ‘possible peatland region’, we hypothesized that continental and dry climates are less favourable to peat formation, so that present circumarctic peatland extent is related to continentality. We supported this hypothesis by comparing a map⁶⁵ of the Gorczynski continentality index⁶⁶ (K_G), based on Climatic Research Unit (CRU) climate data from 1951 to 2000, with maps of peatland extent in flat terrain (ruggedness less than 2) within the current permafrost regions of North America and Eurasia²⁰ ($R^2 = 0.40$; $P = 0.07$; Extended Data Fig. 5). We thus assume that the modern peatland extent is a reasonable analogue of LGM conditions. With a dry, cold climate during the LGM^{67,68}, similar to conditions in highly continental areas today, we reconstruct peatland coverage in the ‘possible peatland regions’ as around 5%, which is the average coverage in the most continental region in Siberia²⁰. In addition, peaty soils may have been present across larger areas, but these are included in the mineral-soil transfer functions from modern analogues to the LGM (see Extended Data Table 2). We assigned areas outside the peatland region a peatland coverage of 1%, so as to not entirely discount peatland presence in localized settings.

Soil carbon-transfer functions. To calculate LGM soil carbon for the different biome and landscape types, we relied on modern-day analogues and the carbon storage in these systems. For the tundra (–steppe) and forest (–steppe) mega biomes, we constructed carbon-transfer functions by extracting soil carbon data from the North American continent presented in the NCSCDV2 database²⁰, which we subdivided into tundra, alpine tundra, taiga and alpine taiga biomes. The biome subdivision was based on the Terrestrial Ecoregions of the World dataset¹⁹. Using the permafrost map of ref. 69, we also categorized these data according to continuous or discontinuous permafrost (including all non-continuous permafrost zones in the discontinuous category).

We decided to use only North American data to calculate our carbon-transfer functions because the spatial soil carbon scaling in this region is explicitly linked to different soil series (US) or soil names (Canada). For other regions, the NCSCDV2 database was created on the basis of more generalized scaling. Where NCSCDV2 is scaled at the soil-series/soil-names level, it has a more realistic representation of landscape scale variability. This in turn translates into a more realistic estimate of scaling errors. There was a concern that simplified thematic scaling, as applied in other NCSCDV2 regions, could cause underestimation of actual variability and associated scaling errors.

For each category, we calculated averages of kg C m^{-2} (transfer functions) normalized by polygon area for each separate category, as follows:

$$\bar{C} = \frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n a_i}$$

where \bar{C} is the weighted average kg C m^{-2} (transfer function), C_i is the mineral carbon content of each polygon belonging to that category (excluding histosols and histels), and a_i is the mineral-soils area for each polygon.

We used relationships of soil carbon with depth for the 1–3-m interval according to an analysis of soil profiles⁷⁰, from which we estimate carbon content at depth as a simple function of carbon content at 0–1 m. As an example, in tundra (–steppe) on continuous permafrost, the carbon content at 2 m depth was 48% of the carbon content at 1 m. For each separate transfer function, we calculated the following:

$$\begin{aligned}\bar{C}_2 &= f(\bar{C}_1) \\ \bar{C}_3 &= f(\bar{C}_1)\end{aligned}$$

where \bar{C}_2 is the weighted average C m^{-2} at 2 m depth, and \bar{C}_3 is for 3 m. Detailed results are given in Extended Data Table 2, with a quick overview in Extended Data Fig. 6.

We estimated transfer functions for steppe (–desert) by using modern-day data from the Qinghai–Tibetan Plateau²¹. The overall means for moist and dry Qinghai–Tibetan Plateau permafrost grasslands were used as analogues for all LGM steppe biomes. Data for the transfer functions for 0–1 m, 1–2 m and 2–3 m were extracted from Fig. 4 and the supplement of ref. 21. Data for 0–30 cm depth were interpolated from a linear regression of $\log(\text{depth})$ to $\log(\text{soil C})$ ($R^2 > 0.99$; $P < 0.05$).

Steep areas in alpine and mixed alpine regions with a slope of more than 4 degrees were treated separately and given a default value of 3 kg C m^{-2} on the basis of ref. 37.

The carbon-transfer functions for peat soils were based on the North American data within NCSCDV2, across all categories regardless of biome, but with distinctions between continuous and discontinuous permafrost as well as between lowland and alpine conditions. These carbon-transfer functions were applied down to 1 m only, because of limited evidence for deeper peat deposits at LGM times. With a few exceptions^{42,43}, most records of LGM peat refer to thin peat layers^{44–49}. Therefore, a carbon-transfer function considering 1 m of peat might still be an overestimate. For 1 m to 3 m, we applied the mineral-soil carbon-transfer functions that corresponded to the assigned biome for that area.

Modern-day soil carbon estimates. The LGM permafrost region extends over the present-day northern permafrost region and over large areas that are presently permafrost free. Modern-day soil carbon stocks for the present northern permafrost region were derived from NCSCDV2²⁰ and from data for the Qinghai–Tibetan Plateau²¹. For areas outside the permafrost region, present-day soil carbon stocks were computed and extracted from the global-scale WISE30sec database⁷¹. This database contains data for the top 2 m of soil. Soil carbon stocks in the 2–3-m depth interval were extrapolated on the basis of biome-specific ratios of soil carbon in the 1–2-m depth interval to the 2–3-m depth interval from table 3 of ref. 72. The spatial scaling of these ratios was applied using spatial biome delineations from ref. 19. Ref. 72 presents a depth distribution for mineral-soil types only. For peatlands (Histosols), soil carbon in the 2–3-m depth range was estimated to be half of the soil carbon content mapped at the 1–2 m. This scaling is consistent with an overall mean peat depth of 2.3 m (ref. 73) and assumptions of a typical mineral-soil carbon content below that. We calculated the uncertainty ranges for these soil carbon estimates by using standard formulas of additive error propagation, combining the uncertainty ranges of the 0–2-m carbon stocks⁷¹ with the uncertainty ranges of the extrapolation ratios⁷². Modern carbon stocks are detailed in Supplementary Table 4.

Lakes. We estimated organic carbon storage in the sediments of large lakes (bigger than 10 km^2) of the northern permafrost region during LGM times and the corresponding area at present⁷⁴, using limited available geochemical data (measured/inferred dry bulk density and organic carbon content) and weighing carbon densities by lake size and their sediment depths. The LGM lake extent was based on lakes reconstructed for LGM times⁷⁵. We estimated carbon storage for large lakes only because the databases used to calculate average soil carbon stocks (NCSCDV2 and WISE30sec) do not spatially resolve small lakes. Therefore they are already included in the soil carbon-transfer functions.

During the LGM, large lakes were limited in extent (occupying around 0.2% of the total LGM permafrost area) and largely restricted to ice-free parts of the Eurasian sector. Storage in sediments from marine isotope stage 4 (MIS4) to the LGM was on average 21 kg C m^{-2} (refs 76–78), resulting in a total stock of 2 Pg C . In postglacial times, the area occupied by large lakes increased (to around 1.9% of the total area), particularly in North America following the retreat of the Laurentide

Ice sheet. Storage in post-LGM sediments is on average 43 kg C m^{-2} (refs ^{76–80}), to which should be added the storage in MIS4–LGM deposits in the Eurasian sector described above. This results in a total present-day estimate for the large-lake area of 45 Pg C . All of these sediments, during both LGM and present-day times, are considered to be in a thawed state and not to contribute to the inert permafrost carbon stocks. Similar to the calculations for thawed-out loess deposits, we consider that carbon stocks in lake sediments pre-dating the MIS4 stage are not part of the LGM or current active carbon cycles.

Phytomass. We calculated the carbon content of phytomass both during the LGM and at present within the LGM northern permafrost region. These estimates are presented in Fig. 2 for reference to the soil carbon storage. The LGM phytomass (30 Pg C) is based on our biome classification together with phytomass estimates⁹. Present-day above-ground and below-ground biomass in the former LGM permafrost and ice-sheet regions was quantified to 61 Pg C (37 Pg C in the LGM permafrost region and 24 Pg C in the ice-sheet extent⁸¹).

Loess reconstruction. We conducted a review of loess and Yedoma studies to calculate the areal extent (Extended Data Fig. 2), average depth and carbon content of those loess sections that lie outside the current permafrost region (Supplementary Table 3). We estimated a total area of additional loess in the LGM northern continuous permafrost region of 2.7 million km^2 , mostly in lowland areas. We calculated average depths of these loess deposits across five separate regional sectors (III, Alaska; IV, northern Europe including northwest Russia; V, Siberia; VI, central loess plateau China; and VII, northeast China). To account for differences in the geochemistry and permafrost extent of each sector, we separated the loess into time intervals of origin of $71\text{--}45 \text{ kyr ago}$ and $45\text{--}19 \text{ kyr ago}$. To avoid double accounting, we removed 2 m from the top (roughly corresponding to an original 3 m of sediment if we account for initial excess ground-ice content), because this interval is accounted for in our estimates of soil organic carbon storage for the $0\text{--}3 \text{ m}$ interval.

To calculate the additional carbon at the time of the LGM, we first conducted an assessment of the present-day carbon content of loess. To infer the carbon densities in these loess deposits at the time of the LGM, we used a survey of published analogues from the present-day Yedoma deposits (Supplementary Tables 2 and 3).

We assume that carbon storage in the current Yedoma region is largely the same as in LGM times (Supplementary Table 3), because the initial carbon losses in Yedoma that resulted from thermokarst following postglacial warming have been compensated by later accumulation in organic-rich lake deposits, peat(y) layers and Holocene soils. Yedoma is also thought to have been prevalent on the Siberian shelf, and therefore we included a deep carbon stock that does not change between past and present. We calculated this carbon stock on the basis of our Yedoma estimates.

A longer and more detailed description of our loess review is available in Supplementary Information.

Glacial burial. We assume that during the LGM, subglacial soil carbon may have been preserved beneath cold-based ice sheets (which are immobile against the ground surface), but that no soil carbon would have been preserved under actively eroding warm-based ice sheets. Following delineations of cold-based ice sheets from ref. ²⁴, we constrained those regions in which we assume that buried permafrost may have been located beneath the Laurentide and Fennoscandian ice sheets⁸² (Extended Data Fig. 7). Assuming that, during glaciation, the areas proximal to expanding ice sheets were tundra environments, we applied a transfer function representative of the high arctic tundra (27.5 kg C m^{-2} down to 3 m ; ref. ¹⁰) across all areas with cold-based ice-sheets and glaciers. We applied this transfer function both on land and on sea shelves down to 3 m . We assumed that peat covered 1% of the area, but, as previously explained, we did not include any peat deeper than 1 m . Below this peat, we applied mineral-soil carbon estimates down to 3 m . Steep alpine regions were again given a value of 3 kg C m^{-2} . We added an estimate for the Greenland Ice Sheet⁶⁹ (1.7 million km^2 , 48 Pg C), using the same procedure as described above, although we conclude that the storage beneath the Greenland Ice Sheet has not changed substantially over time.

Shelf areas beneath ice sheets were estimated using a $\sim 130 \text{ m}$ cut-off⁸³ on the global relief model ETOPO1 (ref. ⁸⁴), meaning that all areas shallower than $\sim 130 \text{ m}$ were included. This is probably an overestimation, as the sea level reached $\sim 130 \text{ m}$ only at the very last stages of the glacial period, when the ice sheets were at their largest configuration.

The total storage of carbon beneath cold-based ice, both on land and on sea shelves, amounts to 123 Pg C . However, if we account for the same potential carbon storage beneath warm-based ice sheets as for cold-based ice sheets, the results show an additional 364 Pg C (Extended Data Table 3).

Inert carbon. We define inert carbon as organic carbon in soils or sediments that is protected from potential mineralization by permafrost. Inert carbon would then slowly be depleted in $\Delta^{14}\text{C}$, and preserve its isotopic signatures of $\delta^{13}\text{C}$ until thaw. Post-thaw microbial processing would also affect the $\delta^{13}\text{C}$ of soil organic matter. We categorize all carbon beneath the active layer as inert, and set the active layer to 30 cm depth across all permafrost soils, following ref. ¹⁰, and consistent with present-day active-layer depths in tundra on North–Central Siberia⁸⁵. For discontinuous permafrost, we calculate 50% of the area to be inert beneath 30 cm depth, while the remaining area is categorized as entirely active rather than inert. To deal with the potential uncertainty in estimates of the permafrost-inert fraction from the discontinuous permafrost extent and active-layer depths, the reported error ranges of the inert fraction include sensitivity analyses (see below). We assume that all deep carbon stocks, as for those within loess, Yedoma and deltas, were inert during the LGM. Carbon preserved beneath cold-based ice sheets is also inert in this scheme.

For the highly uncertain estimate of carbon on the sea shelves, we consider the $0\text{--}3 \text{ m}$ stock to have been disturbed during the deglaciation, removing this stock from the inert carbon storage. For the Yedoma on the Siberian shelf, we assume that a portion equivalent to the loss of inert carbon per area from Yedoma on land has become active beneath the sea floor since the LGM. This loss might be underestimated, as wave erosion may have disturbed the Yedoma ice complexes when the sea advanced onto the shelves.

Scenarios and error estimates. Owing to limitations in NCSCDV2, we are unable to use a standard deviation of our carbon-transfer functions for tundra (–steppe) and forest (–steppe) biomes (see Supplementary Information for details). Instead, we calculate and report ranges of potential minimum-to-maximum LGM carbon stocks for the $0\text{--}3 \text{ m}$ soil. This scheme also deals with area uncertainty within these biome reconstructions. In the first minimum scenario, the tundra (–steppe) and forest (–steppe) categories, both lowland and alpine, were represented by our lowest carbon-transfer function that describes the average carbon content of steppe (–desert). For the steppe (–desert) areas, both lowland and alpine, we calculated a minimum carbon estimate on the basis of the error margins in ref. ²¹.

For this minimum carbon estimate, we applied a peatland extent of 1% across the LGM permafrost landscape. In the maximum scenario, we applied our highest carbon-transfer function, continuous tundra (–steppe), for those regions that were categorized as lowland or alpine tundra (–steppe) or forest (–steppe). We used the carbon-transfer function for continuous forest (–steppe) to the steppe (–desert) biome so as to not underestimate uncertainty. We also applied a peatland coverage similar to that of today (11%)²⁰ across the landscape, but with peatland depth limited to 1 m . Steep slopes were not included in these calculations. These scenarios should fully encompass all uncertainties discussed in Supplementary Information, and are indeed a minimum and a maximum range rather than a range that describes likelihood. We maintain that our best estimate is the most realistic. The error margins for subglacial carbon were calculated by using a $\pm 50\%$ areal coverage of cold-based ice.

Uncertainties in the calculations of the additional LGM permafrost carbon stock in loess deposits are related to area, depth, dry bulk density and per cent carbon estimates. Only for depth have we a nearly complete and consistent dataset (see Supplementary Information). We used standard deviations in reported mean depth for all sectors and the two time periods considered (Supplementary Table 3), to obtain a range of $56\text{--}725$ (366^{+359}_{-310}) Pg C for the LGM stocks and $9\text{--}92$ (48^{+44}_{-39}) Pg C for the present remaining stocks in loess deposits.

We conducted an additional sensitivity analysis of the inert carbon by varying the depth of the active layer ($30\text{--}100 \text{ cm}$) throughout the entire LGM permafrost region and the coverage of permafrost ($10\%\text{--}90\%$ coverage) in our LGM discontinuous permafrost zone. This analysis also meant that we estimated additional carbon in loess for the discontinuous zone (51 Pg C) as a maximum scenario (Supplementary Table 4).

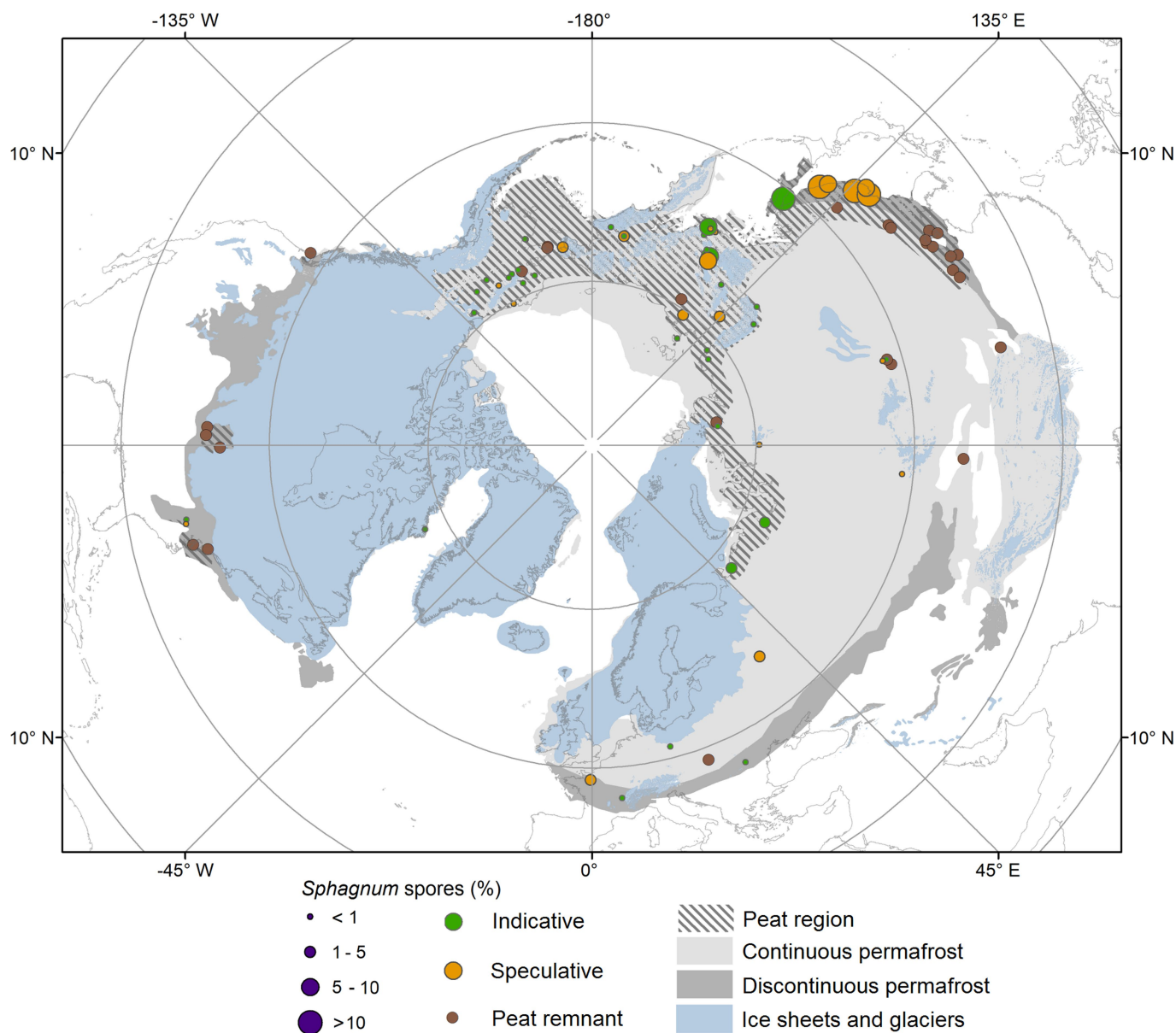
All errors or ranges for individual categories (Table 1 and Supplementary Table 4) have been combined by additive error propagation.

Software. We used ArcMap 10.1 in all geographical computations and MS Excel for the final numerical calculations.

Data availability. The biome reconstruction that supports the findings of this study is available at <https://bolin.su.se/data/Lindgren-2018>, both as a shapefile and in gridded format. Additional sources of used, but unaltered, datasets are referenced within the paper. Compiled datasets are available upon request from the corresponding author.

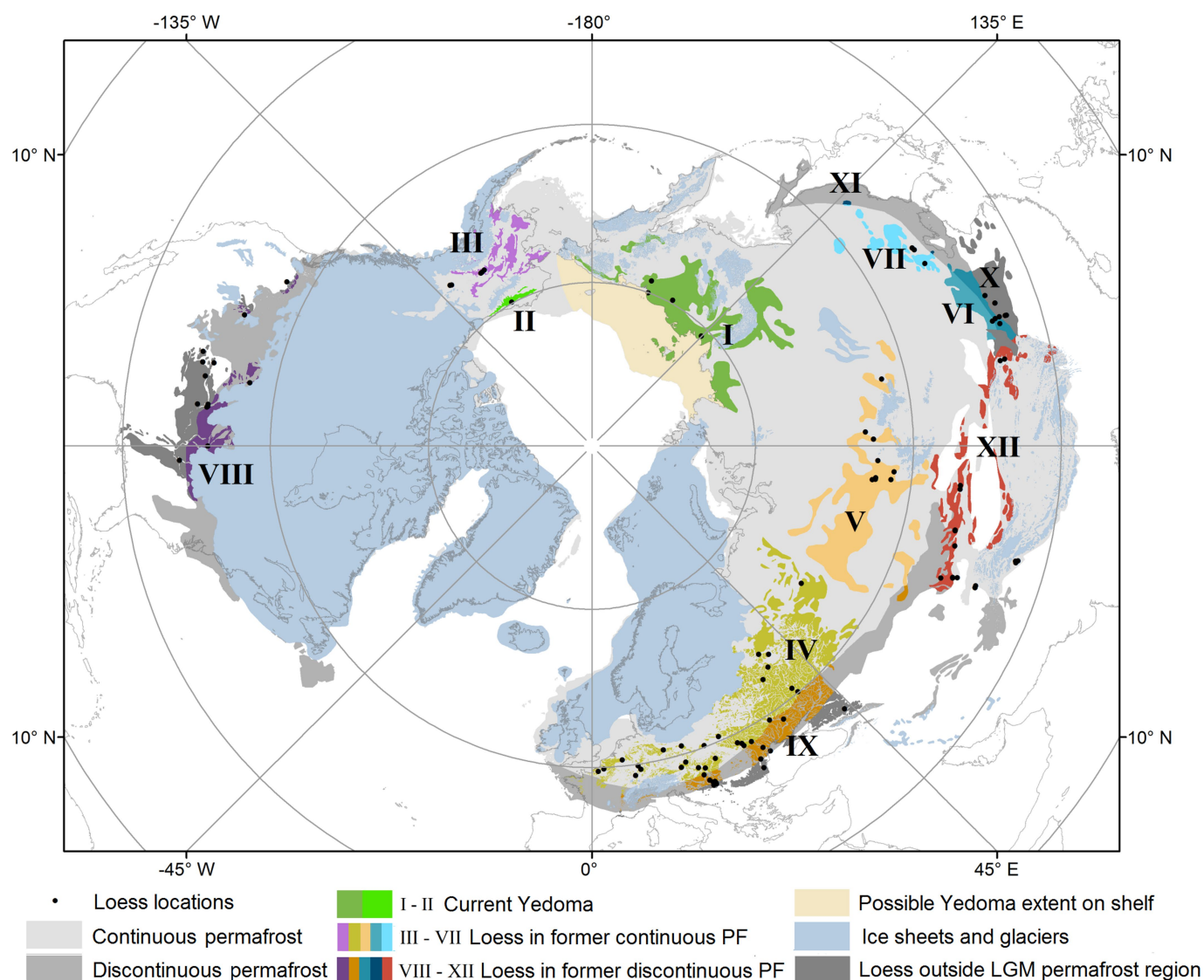
- Lorenzen, E. D. et al. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364 (2011).
- Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P. & Araújo, M. B. Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol.* **6**, e79 (2008).
- Tarasov, P. E. et al. Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *J. Biogeogr.* **27**, 609–620 (2000).
- Andreev, A. A. et al. Paleoenvironmental changes in Northeastern Siberia during the Late Quaternary—evidence from pollen records of the Bykovsky Peninsula. *Polarforschung* **70**, 13–25 (2000).
- Andreev, A. A. et al. Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. *Quat. Sci. Rev.* **30**, 2182–2199 (2011).
- Ni, J., Cao, X., Jeltsch, F. & Herzschuh, U. Biome distribution over the last 22,000yr in China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **409**, 33–47 (2014).

37. Fuchs, M., Kuhry, P. & Hugelius, G. Low below-ground organic carbon storage in a subarctic Alpine permafrost environment. *Cryosphere* **9**, 427–438 (2015).
38. GMTED2010. The Global Multi-resolution Terrain Elevation Data 2010. *US Geological Survey (USGS) Earth Resources Observation & Science (EROS) Center* <https://topotools.cr.usgs.gov> (2011).
39. Dalrymple, J. B., Blong, R. J. & Conacher, A. J. A hypothetical nine unit landsurface model. *Z. Geomorphol.* **12**, 60–76 (1968).
40. Halsey, L. A., Vitt, D. H. & Gignac, L. D. Sphagnum-dominated peatlands in North America since the Last Glacial Maximum: their occurrence and extent. *Bryologist* **103**, 334–352 (2000).
41. Gajewski, K., Viau, A., Sawada, M., Atkinson, D. & Wilson, S. *Sphagnum* peatland distribution in North America and Eurasia during the past 21,000 years. *Glob. Biogeochem. Cycles* **15**, 297–310 (2001).
42. Baker, R. G., Sullivan, A. E., Hallber, G. R. & Horton, D. G. Vegetational changes in Western Illinois during the onset of Late Wisconsinan glaciation. *Ecology* **70**, 1363–1376 (1989).
43. Magyari, E., Jakab, G., Rudner, E. & Sümegi, P. Palynological and plant macrofossil data on Late Pleistocene short-term climatic oscillations in North-Eastern Hungary. *Acta Palaeobot.* **2** (Suppl.), 491–502 (1999).
44. Heusser, C. J. Palynology and phytogeographical significance of a late-Pleistocene refugium near Kalaloch, Washington. *Quat. Res.* **2**, 189–201 (1972).
45. Baker, R. G. et al. A full-glacial biota from southeastern Iowa, USA. *J. Quat. Sci.* **1**, 91–107 (1986).
46. Baker, R. G., Bettis, E. A. I. & Horton, D. G. Late Wisconsinan-early Holocene riparian paleoenvironment in southeastern Iowa. *Geol. Soc. Am. Bull.* **105**, 206–212 (1993).
47. Elias, S. A., Short, S. K., Nelson, C. H. & Birks, H. H. Life and times of the Bering land bridge. *Nature* **382**, 60–63 (1996).
48. Lowery, D., Wah, J. & Rick, T. Post-Last Glacial Maximum dune sequence for the 'Parsonburg' Formation at Elliot's Island, Maryland. *Curr. Res. Pleistocene* **28**, 103–104 (2011).
49. Zhao, Y. et al. Peatland initiation and carbon accumulation in China over the last 50,000 years. *Earth Sci. Rev.* **128**, 139–146 (2014).
50. Andreev, A. A. et al. Late Pleistocene and Holocene vegetation and climate on the Taymyr Lowland, northern Siberia. *Quat. Res.* **57**, 138–150 (2002).
51. Paus, A., Svendsen, J. I. & Matisoff, K. A. Late Weichselian (Valdaian) and Holocene vegetation and environmental history of the northern Timan Ridge, European Arctic Russia. *Quat. Sci. Rev.* **22**, 2285–2302 (2003).
52. de Beaulieu, J.-L. & Reille, M. A long Upper Pleistocene pollen record from Les Echets, near Lyon, France. *Boreas* **13**, 111–132 (1984).
53. Magyari, E. K. et al. Vegetation and environmental responses to climate forcing during the Last Glacial Maximum and deglaciation in the East Carpathians: attenuated response to maximum cooling and increased biomass burning. *Quat. Sci. Rev.* **106**, 278–298 (2014).
54. Bos, J. A. A., Bohncke, S. J. P., Kasse, C. & Vandenbergh, J. Vegetation and climate during the Weichselian early glacial and pleniglacial in the Niederlausitz, eastern Germany—macrofossil and pollen evidence. *J. Quat. Sci.* **16**, 269–289 (2001).
55. Müller, S. et al. Late Quaternary vegetation and environments in the Verkhoyansk Mountains region (NE Asia) reconstructed from a 50-kyr fossil pollen record from Lake Billyakh. *Quat. Sci. Rev.* **29**, 2071–2086 (2010).
56. Demske, D. et al. Late glacial and Holocene vegetation and regional climate variability evidenced in high-resolution pollen records from Lake Baikal. *Global Planet. Change* **46**, 255–279 (2005).
57. Igarashi, Y. & Zharov, A. E. Climate and vegetation change during the late Pleistocene and early Holocene in Sakhalin and Hokkaido, northeast Asia. *Quat. Int.* **237**, 24–31 (2011).
58. Svendsen, J. I. et al. Glacial and vegetation history of the Polar Ural Mountains in northern Russia during the Last Ice Age, Marine Isotope Stages 5–2. *Quat. Sci. Rev.* **92**, 409–428 (2014).
59. Zech, M. et al. Quaternary vegetation changes derived from a loess-like permafrost palaeosol sequence in northeast Siberia using alkane biomarker and pollen analyses. *Boreas* **39**, 540–550 (2010).
60. Wetterich, S. et al. Last Glacial Maximum records in permafrost of the East Siberian Arctic. *Quat. Sci. Rev.* **30**, 3139–3151 (2011).
61. Wetterich, S. et al. Palaeoenvironmental dynamics inferred from late Quaternary permafrost deposits on Kurungnakh Island, Lena Delta, Northeast Siberia, Russia. *Quat. Sci. Rev.* **27**, 1523–1540 (2008).
62. Shichi, K. et al. Vegetation response in the southern Lake Baikal region to abrupt climate events over the past 33 cal kyr. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **375**, 70–82 (2013).
63. Bolikhovskaya, N. S. & Shunkov, M. V. Pleistocene environments of Northwestern Altai: vegetation and climate. *Archaeol. Ethnol. Anthropol. Eurasia* **42**, 2–17 (2014).
64. Neotoma Paleoecology Database. Neotoma, <https://www.neotomadb.org/> (2016).
65. Grieser, J., Gommers, R., Cofield, S. & Barandi, M. Data sources for FAO worldmaps of Koeppen climatologies and climatic net primary production. *Food and Agriculture Organization of the United Nations* http://www.fao.org/nr/climpag/globgrids/KC_commonddata_en.asp (2006).
66. Gorczynski, W. Sur le calcul du degré de continentalisme et son application dans la climatologie. *Geogr. Ann.* **2**, 324–331 (1920).
67. Scheff, J., Seager, R., Liu, H. & Coats, S. Are glacials dry? Consequences for paleoclimatology and for greenhouse warming. *J. Clim.* **30**, 6593–6609 (2017).
68. Bartlein, P. J. et al. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Clim. Dyn.* **37**, 775–802 (2011).
69. Brown, J., Ferrains, O., Heginbottom, J. & Melnikov, E. Circum-Arctic map of permafrost and ground-ice conditions, version 2. *National Snow and Ice Data Center* <https://nsidc.org/data/ggd318> (2002).
70. Hugelius, G. et al. A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. *Earth Syst. Sci. Data* **5**, 393–402 (2013).
71. Batjes, N. H. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**, 61–68 (2016).
72. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
73. Gorham, E. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecol. Appl.* **1**, 182–195 (1991).
74. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1–22 (2004).
75. Amosov, M. Lake-levels, vegetation and climate in Central Asia during the Last Glacial Maximum. *Russ. Geogr. Soc. Her. (in Russian)* **910–911**, 1–14 (2014).
76. Ebel, T., Melles, M. & Niessen, F. in *Land-Ocean Systems in the Siberian Arctic: Dynamics and History* (eds Kassens, H. et al.) 425–435 (Springer, Berlin, 1999).
77. Colman, S. M., Carter, S. J., Hatton, J. & Haskell, B. J. *Cores collected in Lake Baikal, Siberia, by the U.S. Geological Survey, 1990 to 1992: Visual Descriptions, Photographs, X-radiographs, Bulk-Density Measurements, and Grain-Size Analyses* (US Dept Interior Geol. Surv. Open-File Report 94-445, 1992).
78. Orem, W. H., Colman, S. M. & Lerch, H. E. Lignin phenols in sediments of Lake Baikal, Siberia: application to paleoenvironmental studies. *Org. Geochem.* **27**, 153–172 (1997).
79. Teller, J. T. & Last, W. M. Late Quaternary History of Lake Manitoba, Canada. *Quat. Int.* **16**, 97–116 (1981).
80. Lim, J., Woodward, J., Tulaczyk, S., Christoffersen, P. & Cummings, S. P. Analysis of the microbial community and geochemistry of a sediment core from Great Slave Lake, Canada. *Antonie van Leeuwenhoek* **99**, 423–430 (2011).
81. Ruesch, A. & Gibbs, H. K. New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000. *Carbon Dioxide Information Analysis Center* <http://cdiac.ess-dive.lbl.gov> (2008).
82. Ehlers, J., Gibbard, P. & Hughes, P. *Quaternary Glaciations—Extent and Chronology: A Closer Look* (Elsevier, London, 2011).
83. Clark, P. U. et al. The Last Glacial Maximum. *Science* **325**, 710–714 (2009).
84. Amante, C. & Eakins, B. W. *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis* (NOAA Technical Memorandum NESDIS-NGDC-24, Boulder, Colorado, 2009).
85. Siewert, M. B., Hugelius, G., Heim, B. & Faucherre, S. Landscape controls and vertical variability of soil organic carbon storage in permafrost-affected soils of the Lena River Delta. *Catena* **147**, 725–741 (2016).
86. Lewis, G., Fosberg, M., McDole, R. & Chugg, J. Distribution and some properties of loess in south-central and south-eastern Idaho. *Soil Sci. Soc. Am. Proc.* **39**, 1165–1168 (1975).
87. Romanovsky, N. N. *Fundamentals of Cryogenics of Lithosphere* (Moscow Univ. Press, Moscow, 1993).
88. Busacca, A. & McDonald, E. Regional sedimentation of late Quaternary loess on the Columbia Plateau: sediment source areas and loess distribution patterns, regional geology of Washington State. *Washingt. Div. Geol. Earth Resour. Bull.* **80**, 181–190 (1994).
89. Bettis, E. A., Muhs, D. R., Roberts, H. M. & Wintle, A. G. Last Glacial loess in the conterminous USA. *Quat. Sci. Rev.* **22**, 1907–1946 (2003).
90. Haase, D. et al. Loess in Europe—its spatial distribution based on a European Loess Map, scale 1:2,500,000. *Quat. Sci. Rev.* **26**, 1301–1312 (2007).
91. Muhs, D. R. The geologic records of dust in the quaternary. *Aeolian Res.* **9**, 3–48 (2013).



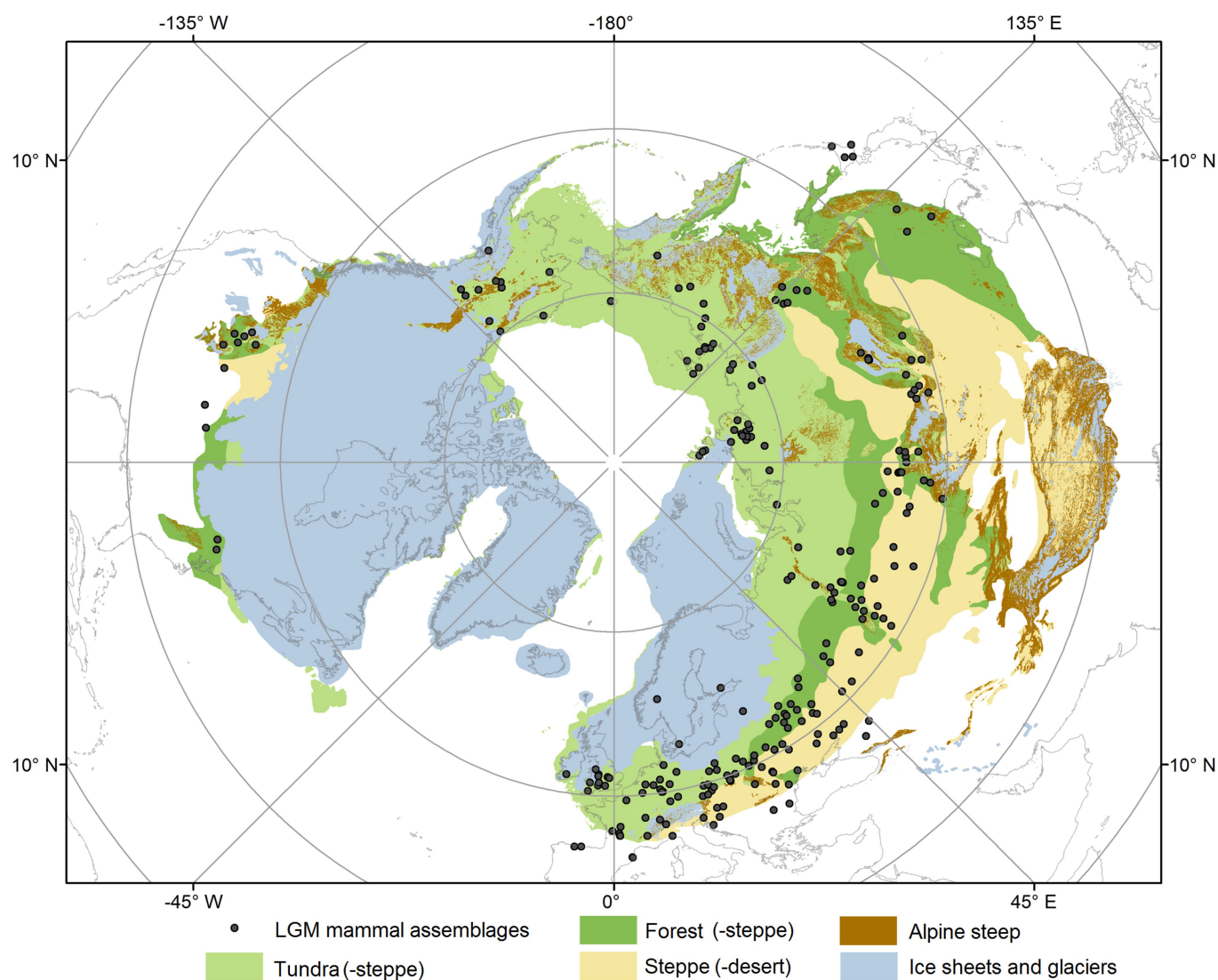
Extended Data Fig. 1 | LGM peat region. The reconstructed peat region is based on already-reconstructed areas^{40,41}, *Sphagnum* spore evidence, and the occurrence of peat^{42,43} or peaty layers^{44–49}. The colouring and size of these points show the percentage of the total pollen sum that was spores (not algae) and our interpretation of the reliability of the dating. Indicative

ages are better constrained than speculative ages (see Supplementary Information). Evidence of dated peat or peaty deposits is shown in dark brown. Data for ice sheets and glaciers are modified from ref. ⁸², and the permafrost region¹¹ is included for reference.



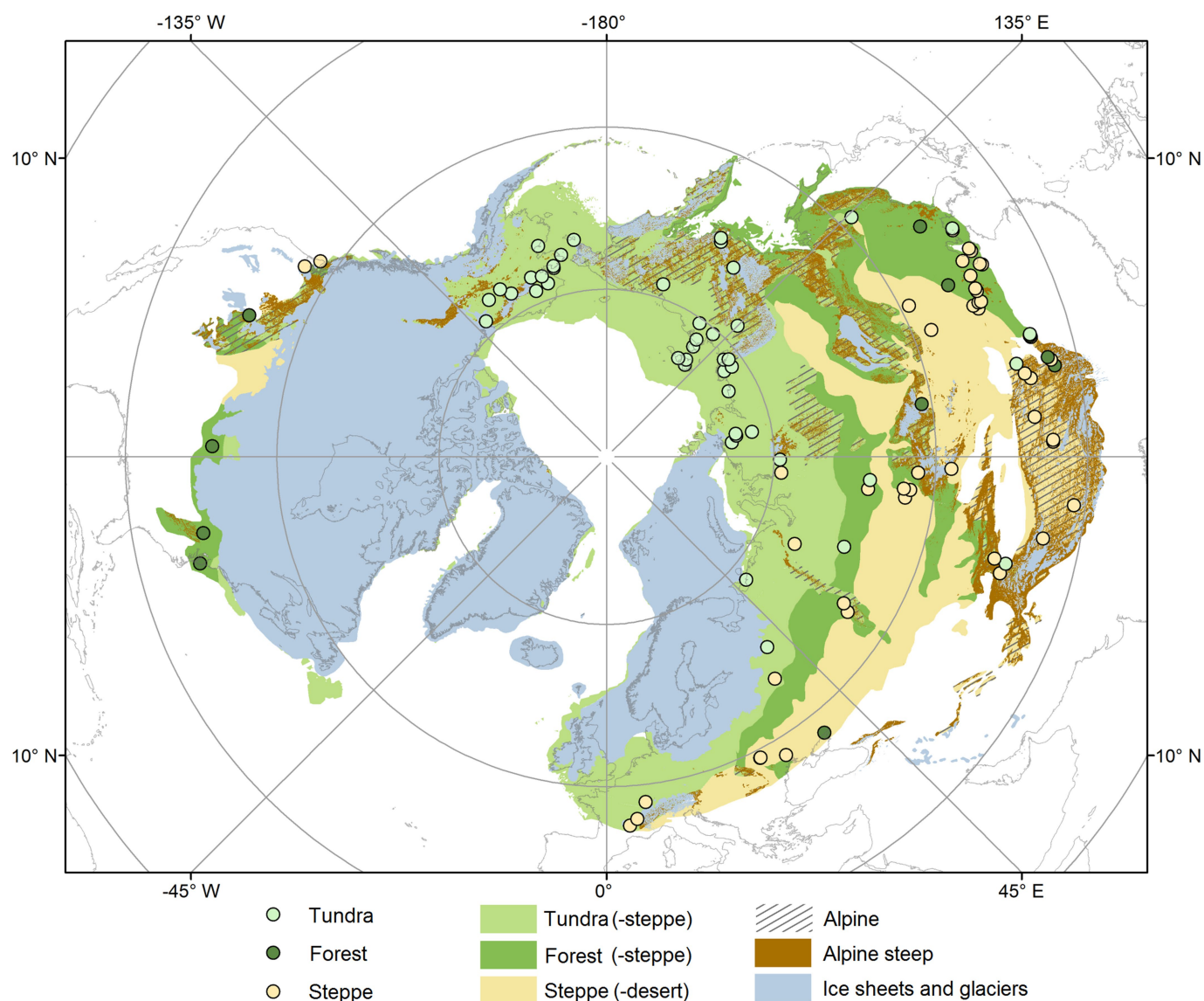
Extended Data Fig. 2 | Loess and Yedoma deposits during the LGM. The deposits were compiled from several data sources^{86–91}, and separated into sections (shown with Roman numerals) as described in Methods and Supplementary Information. The loess extent outside of the LGM continuous permafrost region is included for reference⁸⁹. A tentative

area of Yedoma extent on the shelf is also included. We assume that this area had the same degree of dissection as the Yedoma on land (see Supplementary Information). Data for ice sheets and glaciers are modified from ref. ⁸², and the permafrost region¹¹ is included for reference. PF, permafrost.



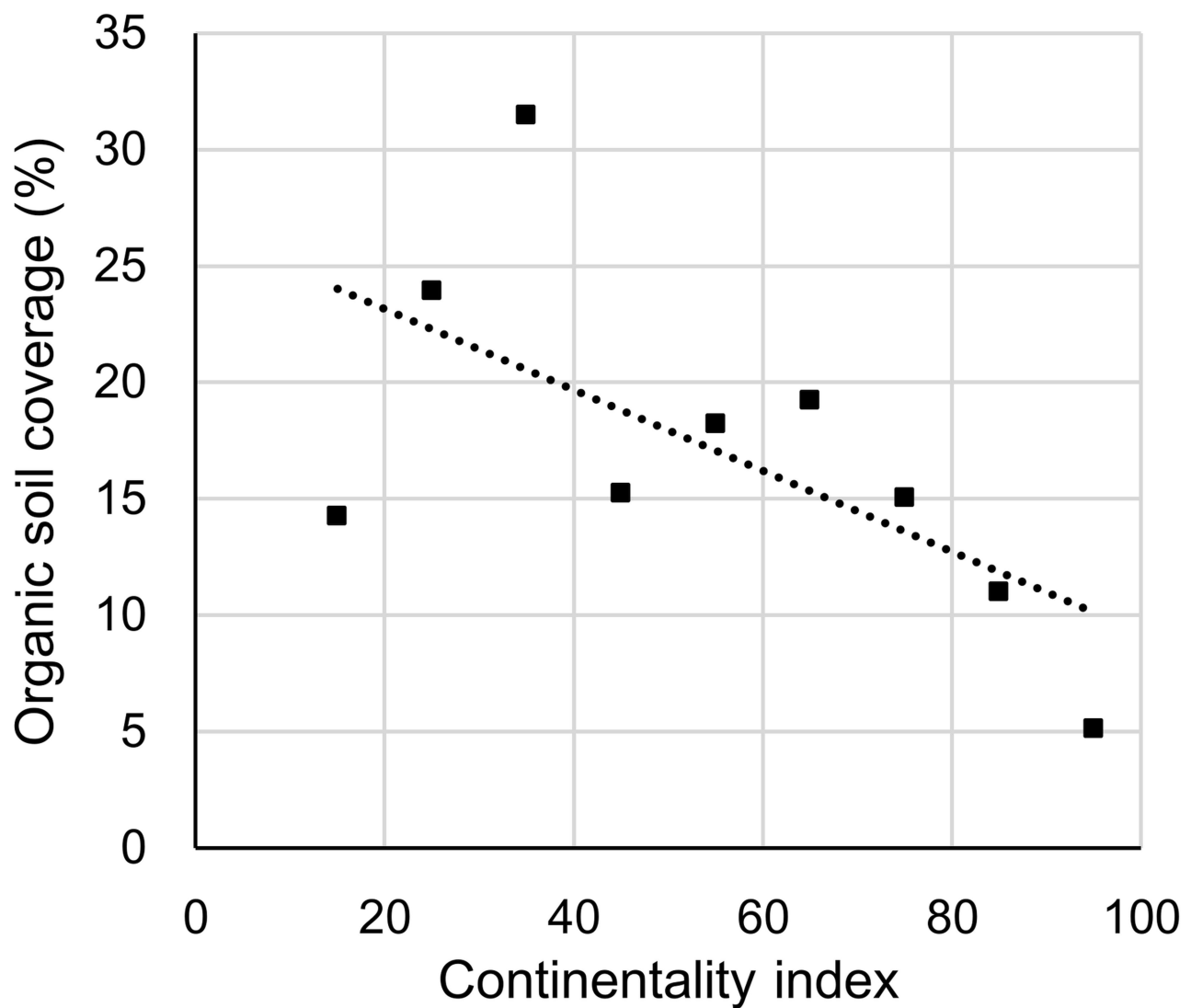
Extended Data Fig. 3 | LGM mega biomes and LGM mammal assemblages. Assemblages of mammoths, horses, bison, reindeer, wholly rhinoceroses and muskoxen^{17,31,32} dated to between 18 kyr BP and 26 kyr BP indicate an environment that was productive enough to support

megafauna (see Supplementary Information). Note that none of the data points within the Fennoscandian Ice Sheet are younger than 19 kyr BP. Data for ice sheets and glaciers are modified from ref.⁸².



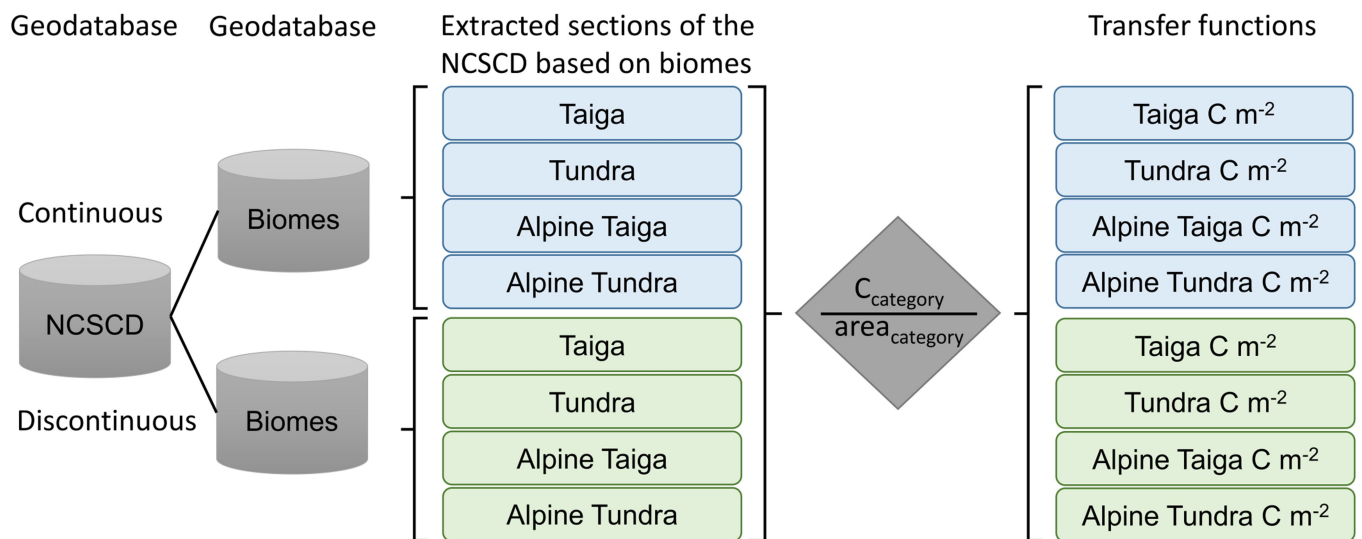
Extended Data Fig. 4 | The LGM mega biomes, and point data from pollen and macrofossil findings. The map shows the major biomes within the LGM permafrost region, constructed from three separate empirical maps^{15–17}, as well as our additional separation of alpine

environments and steep areas¹⁸. Biomized pollen data and macrofossil findings^{9,33–36} were compared with the reconstruction to assess its accuracy (see Supplementary Information). Data for ice sheets and glaciers are modified from ref.⁸².



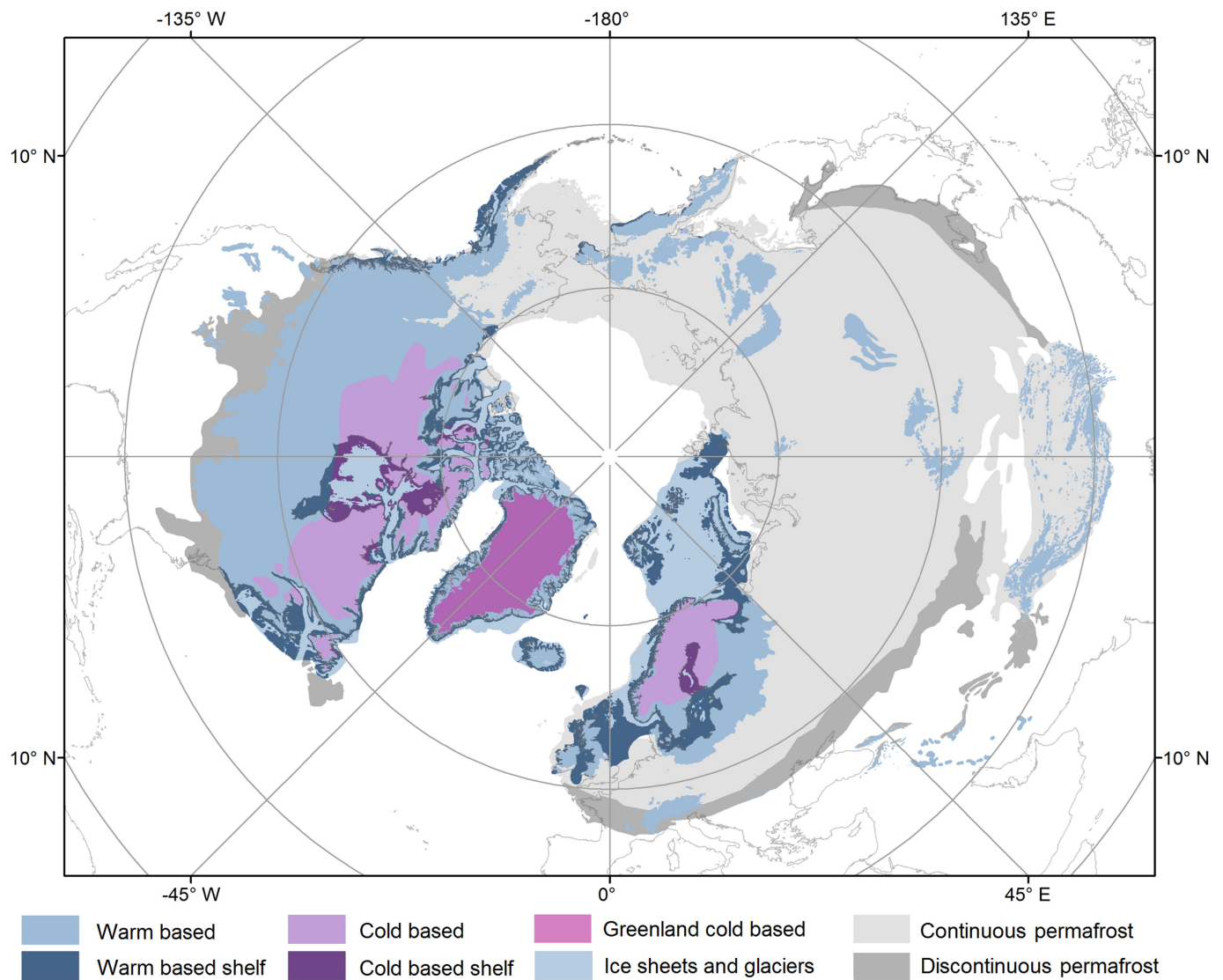
Extended Data Fig. 5 | Regression between continentality and organic soil coverage. The organic soil (peat) coverage was calculated from data within the NCSCDv2 database for flat terrain only (see Supplementary Information). The data were aggregated into classes of continentality,

determined from the map⁶⁵ of the Gorchynski continentality index⁶⁶ (see Methods). The trend indicates that peat coverage in flat terrain is lower in regions with high continentality ($R^2 = 0.4$) than in regions of low continentality.



Extended Data Fig. 6 | A schematic presentation of data handling for soil of depth 0–1 m. To estimate LGM soil carbon, we used databases to calculate carbon-transfer functions for different biomes. The colouring describes the continuous and discontinuous sections that were separated

before applying the biomes as a second filter. Modern-day biomes were overlain with modern-day carbon stocks in permafrost terrain, providing biome-specific information that was translated into transfer functions (see Methods).



Extended Data Fig. 7 | Warm-based and cold-based sections of ice sheets and glaciers^{24,69}, both on land and on shelves. Cold-based areas are assumed to retain the carbon storage formed before the glaciation. Warm-based ice sheets and glaciers, on the other hand, are erosive, and

we assume no preserved carbon storage. Data for ice sheets and glaciers are modified from ref. ⁸², and the permafrost region¹¹ is included for reference.

Extended Data Table 1 | Biome categories included under each ‘mega’ biome for lowland and alpine areas

Lowland		
Tundra (-steppe)	Forest (-steppe)	Steppe (-desert)
Tundra	Coniferous/broad-leaved and montane forest	Mammoth tundra steppe (boreal)
Complex of tundra steppe forest, local halophytic	Open pine forest of low-mountainous regions	Grass steppe and scattered flatland semi desert
Mammoth tundra steppe (arctic)	Appalachian forest	Periglacial steppe dominated by European elements
Subarctic desert, montane tundra, subalpine meadows ¹	Boreal forest	Periglacial steppe dominated by Mongolian elements
	Floridan forest subtropical	Mixed-grass flatland steppe
	Forest tundra	Mongolian desert steppe
	Dark coniferous and birch montane forest	Montane steppe and semi desert
	Forest refugia	Central North American steppe
	Forest steppe with birch and pine	Mixed mammoth tundra-steppe (boreal) w. steppe
	Japan-Chinese forest steppe	Steppe
	Light coniferous montane forest	Pontian-Kazakhstan steppe
	Mixed mammoth tundra-steppe (boreal) w. forest steppe	
	Open birch and spruce forest	
	Open forest – larch and birch with tundra elements	
	Periglacial forest steppe, larch pine birch tundra	
	Subalpine forest*	
	Mixed mammoth tundra steppe (boreal) w. forest	
	Forest steppe with European broadleaf trees	
Alpine		
Tundra (-steppe)	Forest (-steppe)	Steppe (-desert)
Subarctic desert, montane tundra, subalpine meadows	Subalpine forest	Mountain desert
Alpine tundra		Montane steppe and semi desert

*Classified as lowland on shelf areas.

Extended Data Table 2 | Areas, carbon-transfer functions and carbon stocks (in Pg C) at the LGM

Environment	Permafrost	Area km ²	30 cm kg Cm ⁻²	0 - 1m kg Cm ⁻²	1 - 2 m kg Cm ⁻²	2 -3 m kg Cm ⁻²	Pg C
Lowland*							
Steppe (-desert)	continuous	6,424,000	4.2	7.5	3.1	2.8	86
	discontinuous	1,813,000	4.2	7.5	3.1	2.8	24
Forest (-steppe)	continuous	5,085,000	9.4	20.2	10.1	12.1	217
	discontinuous	1,413,000	8.4	14.8	7.4	8.9	45
Tundra (-steppe)	continuous	10,353,000	11.2	26.7	12.8	10.8	527
	discontinuous	635,000	9.2	13.1	6.3	5.3	16
Peat	continuous	416,000	16.9	63.1	-	-	26
	discontinuous	76,000	17.5	62.2	-	-	5
Alpine							
steep >4 deg	N/A	4,084,000	-	3.0	-	-	12
Steppe (-desert)	continuous	1,288,000	4.2	7.5	3.1	2.8	17
	discontinuous	198,000	4.2	7.5	3.1	2.8	3
Forest (-steppe)	continuous	12,000	10.1	13.5	6.7	8.1	0.3
	discontinuous	187,000	5.5	11.7	5.8	7.0	5
Tundra (-steppe)	continuous	2,147,000	5.2	11.4	5.5	4.6	47
	discontinuous	201,000	6.6	13.6	6.5	5.5	5
Peat	continuous	74,000	19.4	67.2	-	-	5
	discontinuous	8,000	14.7	69.4	-	-	1
Glacial burial							
Cold based		4,320,000	-	17.8	6.9	2.8	119
Cold based peat		44,000	-	63.1	-	-	3
Cold based steep		304,000	-	3.0	-	-	1
Greenland		1,713,000		17.8	6.9	2.8	47
Greenland peat		17,000	16.9	63.1	-	-	1
Deep deposits							
Yedoma		1,200,000					301
Loess		2,700,000					366
Yedoma on shelf		1,600,000					349
Delta		76,000					91
Lakes							
Large Lakes		70,000					2

*Shelf areas and resulting carbon stocks are included under the lowland category.

Extended Data Table 3 | Estimates of areas and carbon stocks (in Pg C) beneath ice sheets and glaciers

Environment	Area	0 - 1 m C m ⁻²	1 - 2 m C m ⁻²	2- 3 m C m ⁻²	Pg C
Land					
Cold based	3,773,000	17.8	6.9	2.8	104
Cold based peat	38,000	63.1			2.4
Alpine steep	304,000	3.0			0.9
Warm based (not Greenland)	9,789,000	17.8	6.9	2.8	270
Warm based peat	99,000	63.1			6.2
Alpine steep	2,675,000	3.0			8.0
Shelf					
Cold based	547,000	17.8	6.9	2.8	15
Cold based peat	6,000	63.1			0.3
Alpine steep	0	3.0			0.0
Warm based (not Greenland)	2,821,000	17.8	6.9	2.8	77.8
Warm based peat	28,000	63.1			1.8
Alpine steep	0	3.0			

Xenon isotopic constraints on the history of volatile recycling into the mantle

Rita Parai^{1*} & Sujoy Mukhopadhyay²

The long-term exchange of volatile species (such as water, carbon, nitrogen and the noble gases) between deep Earth and surface reservoirs controls the habitability of the Earth's surface. The present-day volatile budget of the mantle reflects the integrated history of outgassing and retention of primordial volatiles delivered to the planet during accretion, volatile species generated by radiogenic ingrowth and volatiles transported into the mantle from surface reservoirs over time. Variations in the distribution of volatiles between deep Earth and surface reservoirs affect the viscosity, cooling rate and convective stress state of the solid Earth. Accordingly, constraints on the flux of surface volatiles transported into the deep Earth improve our understanding of mantle convection and plate tectonics. However, the history of surface volatile regassing into the mantle is not known. Here we use mantle xenon isotope systematics to constrain the age of initiation of volatile regassing into the deep Earth. Given evidence of prolonged evolution of the xenon isotopic composition of the atmosphere^{1,2}, we find that substantial recycling of atmospheric xenon into the deep Earth could not have occurred before 2.5 billion years ago. Xenon concentrations in downwellings remained low relative to ambient convecting mantle concentrations throughout the Archaean era, and the mantle shifted from a net degassing to a net regassing regime after 2.5 billion years ago. Because xenon is carried into the Earth's interior in hydrous mineral phases^{3–5}, our results indicate that downwellings were drier in the Archaean era relative to the present. Progressive drying of the Archaean mantle would allow slower convection and decreased heat transport out of the mantle, suggesting non-monotonic thermal evolution of the Earth's interior.

Volatiles are degassed from the interior to Earth's surface reservoirs during partial melting of upwelling mantle. Conversely, downwelling mantle transports material from Earth's surface to the deep interior. Some volatiles are removed from downwellings and expelled back to the surface via magmatism, but some may be retained within downwellings and ultimately mixed into the convecting mantle (that is, the mantle source of mid-ocean ridge basalts). Here we use the term 'regassing' to indicate the transport of surface volatiles into the mantle beyond depths of magma generation and subsequent mixing into the convecting mantle. In early Earth history, downwelling return flow may have differed in nature from modern subduction⁶, where surface volatiles are regassed into the mantle in association with hydrothermally altered subducting slabs^{7–10}. The nature of early downwellings and the timing of the onset of substantial volatile regassing are not known.

Xe isotopes provide a powerful tool with which to probe the history of volatile cycling on the Earth. The nine isotopes of Xe provide insight into the integrated history of mantle volatile delivery, degassing and regassing: ¹²⁴Xe, ¹²⁶Xe, ¹²⁸Xe and ¹³⁰Xe are primordial, radiogenic ¹²⁹Xe is produced by decay of short-lived ¹²⁹I, and ¹³¹Xe, ¹³²Xe, ¹³⁴Xe and ¹³⁶Xe are produced in characteristic proportions and on different timescales by the spontaneous fission of short-lived ²⁴⁴Pu (half-life of $t_{1/2} = 80.0$ Myr) and long-lived ²³⁸U ($t_{1/2} = 4.468$ Gyr). Degassing fractionates lithophile parent isotopes from atmophile Xe daughter

isotopes, so that mantle Xe isotopic compositions are sensitive to degassing on a variety of timescales. Critically, mantle Xe is also sensitive to volatile regassing: atmospheric Xe, which is isotopically distinct from mantle Xe, has been regassed to the deep Earth in sufficient quantities to affect mantle Xe isotopic compositions^{11–16}.

The isotopic composition of atmospheric Xe available for regassing has varied over time. Compared to primordial noble gas compositions, the Earth's modern atmosphere is depleted in Xe relative to other noble gases, and atmospheric Xe is isotopically mass-fractionated^{17–19}. These observations are attributed to mass-fractionating loss of Xe from the atmosphere. Xe measured in fluid inclusions in Archaean rocks of different age indicate that this process was protracted and that the modern atmospheric composition was attained about 2.0 Gyr ago^{1,2} (Extended Data Fig. 1). If atmospheric volatiles were regassed into the deep Earth throughout Earth history, then the mantle has received a mixture of modern and ancient atmospheric Xe isotopic compositions over time.

Assuming that all regassed atmospheric Xe is modern, the present-day mantle Xe budget is found to be dominated by regassed atmospheric Xe^{12,14–16} (~80%–90%, consistent with estimates of the proportion of regassed atmospheric Xe from the stable, non-radiogenic isotopes of Xe in continental well gases¹¹). If all regassed Xe instead had the isotopic composition estimated for the atmosphere 3.3 Gyr ago²⁰, then the present-day upper-mantle Xe composition could not be explained (Extended Data Fig. 2). This test of end-member scenarios suggests that the budget of regassed atmospheric Xe retained within the mantle today is likely to be predominantly modern (<2.0 Gyr ago²) in its isotopic composition, consistent with two possible scenarios: (a) early Xe regassing was suppressed, perhaps as high mantle potential temperatures promoted shallow release of atmospheric volatiles from downwelling material in the past, or (b) substantial quantities of ancient (>2.0 Gyr ago) atmospheric Xe were regassed beyond depths of magma generation into the deep Earth, but strong subsequent mantle degassing, in association with high mantle processing rates, depleted the mantle of ancient regassed volatiles. High mantle processing rates would promote both strong degassing and concurrent regassing if Xe were retained in downwellings early in Earth history. Intensive degassing and regassing would diminish the mantle radiogenic (¹²⁹Xe) and fissiogenic (^{131,132,134,136}Xe) excesses relative to the atmosphere and would affect the proportion of ²⁴⁴Pu-derived to ²³⁸U-derived fissiogenic xenon. To distinguish between scenarios (a) and (b), it is thus necessary to explicitly model continuous degassing, regassing and fissiogenic production in the mantle over time.

We use a forward model of mantle Xe transport and ingrowth to explore limits on the history of Xe regassing into the mantle. We apply three model forcings (Fig. 1). (1) We prescribe a mass-fractionating atmospheric Xe isotopic composition over time on the basis of data from Archaean rocks² (Extended Data Fig. 1). (2) The mantle processing-rate history is explored as a free parameter. (3) The concentration of Xe retained in downwellings beyond depths of magma generation over time (¹³⁰Xe_d time series, referred to here as 'regassing history') is

¹Department of Earth and Planetary Sciences, Washington University in St. Louis, Saint Louis, MO, USA. ²Department of Earth and Planetary Sciences, University of California, Davis, Davis, CA, USA. *e-mail: parai@wustl.edu

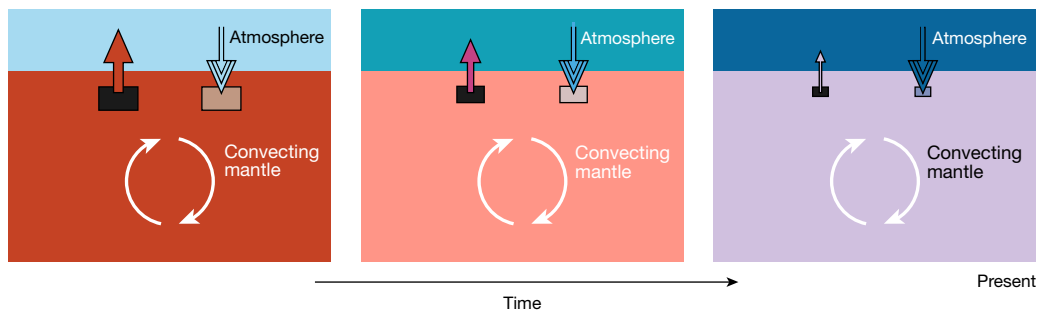


Fig. 1 | Conceptual model of mantle degassing, regassing and fissiogenic production. At a given time step, a parcel of upwelling mantle undergoes partial melting and is completely degassed of its Xe contents (red, pink and purple arrows). An equal mass of downwelling material carries atmospheric Xe to be regassed into the mantle (blue arrows). Convective mixing is indicated by the white arrows. The mantle processing rate decreases over time (as illustrated by the smaller processed-mantle boxes). Many different potential regassing histories ($^{130}\text{Xe}_d$) are tested over time (indicated by the multiple blue arrows). Atmospheric Xe undergoes mass-dependent fractionation over time until 2 Gyr ago (as illustrated by the different shades of blue in the atmosphere box). The net change in ^{130}Xe in the mass of mantle that is processed at a given time step is the

difference between the instantaneous mantle ^{130}Xe concentration and the instantaneous $^{130}\text{Xe}_d$. Assuming instantaneous mixing, the net change is distributed from the processed mantle parcel to the full convecting mantle reservoir. Concentrations of all other Xe isotopes in the mantle are computed using the instantaneous mantle isotopic composition for degassing, the instantaneous atmospheric Xe isotopic composition for regassing, and the instantaneous mantle ^{238}U and ^{244}Pu concentrations to determine fissiogenic ingrowth of $^{131,132,134,136}\text{Xe}$ (Methods). The convecting mantle ^{130}Xe concentration and Xe isotopic composition accordingly evolve over time in response to mantle degassing, regassing of atmospheric Xe with a changing isotopic composition and Xe production by spontaneous fission.

explored as a free parameter set. (A summary of the notation used in the paper is given in Extended Data Table 1.) The initial mantle ^{130}Xe concentration is taken to be consistent with a late-veneer contribution of Xe from carbonaceous chondrites of 0.1%–1% of the Earth's mass²¹, although we note that the primordial Xe budget may have been partially acquired during the main stage of accretion. The initial mantle Xe isotopic composition is taken to be that of average carbonaceous chondrite^{17,18,22}. The Xe isotopic composition of the atmosphere is modelled by a Rayleigh fractionation trajectory towards the modern composition (Extended Data Fig. 1). We search for Xe regassing histories that yield mantle compositions consistent with Xe isotopic and concentration constraints from the literature (Methods, Extended Data Table 2).

The mantle processing-rate history describes the changing mass flux (in grams per unit time) of upwelling mantle that undergoes partial melting at the surface over time. We assume that partial melting results in complete degassing of the processed mantle mass. The degassing Xe flux at any given time is thus determined by the instantaneous processing rate and Xe concentration in the mantle. An equal mass flux of downwelling mantle carries regassed surface Xe into the deep mantle (Fig. 1). The regassing Xe flux at any given time thus reflects the instantaneous processing rate and Xe concentration in downwellings, $^{130}\text{Xe}_d$. We use an exponentially decreasing mantle processing rate pinned at the present-day mid-ocean ridge processing rate. Different processing-rate histories are explored through variation in the exponential time constant (τ ; see Methods). The mantle processing-rate history may be expressed in terms of the number of mantle reservoir masses processed over Earth history (N_{res}). We use values of η that yield whole number values of N_{res} and explore the effect of varying the size of the convecting mantle relative to the whole mantle ($M_{\text{res}} = 50\%–90\%$).

To track fissiogenic Xe ingrowth concurrently with degassing and regassing, U and Pu concentrations must be tracked in the mantle over time. The initial mantle U concentration is taken to be 21 parts per billion (p.p.b.; bulk silicate Earth), and the initial ^{244}Pu concentration is computed assuming an initial $^{244}\text{Pu}/^{238}\text{U}$ ratio based on chondrites (Methods). U and Pu are partially sequestered into the continental crust over time. Both species are highly incompatible; however, the extraction of U and Pu from the mantle during partial melting may be offset by recycling of U- and Pu-rich materials in downwellings at a given time step. To model net extraction, U and Pu loss from the mantle at each time step directly tracks continental crustal growth over time (Extended Data Fig. 3, Methods).

The amount of atmospheric Xe recycled beyond depths of magma generation into the Earth's interior varied over time, as downwelling

lithologies and mantle pressure–temperature conditions evolved on a cooling Earth. We use our model to test many different regassing histories ($^{130}\text{Xe}_d$; Fig. 2). A Monte Carlo numerical approach is used to achieve efficient coverage of a wide regassing history parameter space: for each model realization, a different potential regassing history is

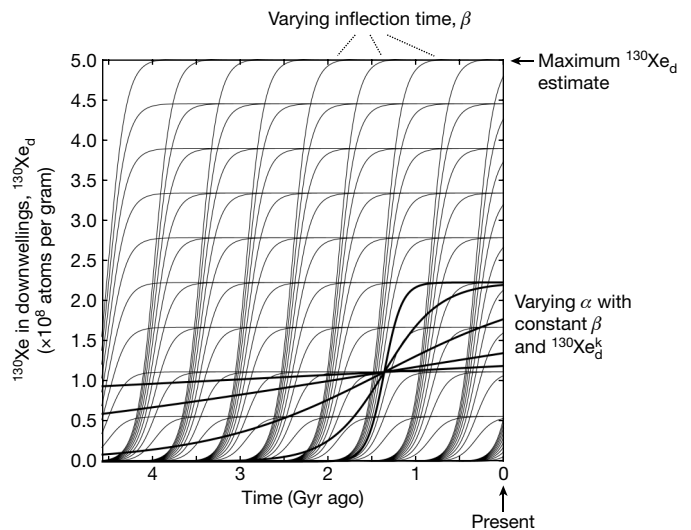


Fig. 2 | Example sigmoidal regassing histories. Three numerical parameters describe the sigmoid (Methods): $^{130}\text{Xe}_d^k$ is the downwelling ^{130}Xe -carrying capacity, α is the growth rate and β is the sigmoid inflection time. A Monte Carlo numerical method is used to explore the parameter space efficiently and test a wide variety of sigmoidal $^{130}\text{Xe}_d$ time series. Light-grey lines represent a collection of sigmoids with constant α and 10 different $^{130}\text{Xe}_d^k$ and β values. Thick solid black lines illustrate the result of varying the growth rate, α , for a single ($^{130}\text{Xe}_d^k$, β) pair. Sampling a limited time interval yields only a portion of the sigmoid shape, such that the initial $^{130}\text{Xe}_d$ values may be greater than 0, and the present-day $^{130}\text{Xe}_d$ values are lower than or equal to the carrying capacity, $^{130}\text{Xe}_d^k$. The growth rate is allowed to vary between 10^{-10} Gyr $^{-1}$ and 10^{-8} Gyr $^{-1}$, with small α corresponding to slow growth. The inflection time is allowed to vary between 0.08 Gyr and 10 Gyr after the formation of the Solar System, and the carrying capacity ranges from 0 to an upper bound of 5×10^8 atoms ^{130}Xe per gram (Methods). Extended Data Fig. 3 shows examples of exponential $^{130}\text{Xe}_d$ time series for comparison. A sigmoidal functional form enables testing of a wide variety of regassing histories, including near-linear, near-exponential and step functions.

determined by randomly drawing values for the parameters that define $^{130}\text{Xe}_d$ as a function of time (Methods, Extended Data Table 1). A diverse assortment of regassing histories are tested using a sigmoidal $^{130}\text{Xe}_d$ functional form (Fig. 2; see Extended Data Fig. 3 for exponential $^{130}\text{Xe}_d$ examples).

Over a given time step, the net change in mantle ^{130}Xe concentration represents the balance between ^{130}Xe lost by degassing of the mantle mass processed by partial melting and ^{130}Xe gained by regassing via a corresponding downwelling mass with concentration $^{130}\text{Xe}_d$ (Fig. 1). Regassing of ^{130}Xe is computed on the basis of $^{130}\text{Xe}_d$ and the instantaneous atmospheric $^{128,131,132,134,136}\text{Xe}/^{130}\text{Xe}$ composition. Fissionogenic production is calculated using the instantaneous concentrations of ^{238}U and ^{244}Pu in the convecting mantle (Methods). The model thus tracks how the ^{130}Xe concentration and Xe isotopic composition in the mantle ($^{128}\text{Xe}/^{130}\text{Xe}$, $^{128}\text{Xe}/^{132}\text{Xe}$, $^{130}\text{Xe}/^{132}\text{Xe}$, $^{131}\text{Xe}/^{132}\text{Xe}$, $^{134}\text{Xe}/^{132}\text{Xe}$ and $^{136}\text{Xe}/^{132}\text{Xe}$) respond to degassing, regassing and fissionogenic production. Xe isotopic evolution paths corresponding to four potential regassing histories illustrate how the mantle Xe isotopic composition changes over time in response to model forcings (Fig. 3). A model realization is successful if two criteria are met: (1) the present-day concentration of ^{130}Xe in the mantle falls within the estimated range of 4.3×10^5 to 9.2×10^5 atoms of ^{130}Xe per gram (Methods) and (2) the present-day mantle Xe isotopic composition falls within the estimated convecting mantle composition field^{11,14,16,23} (Fig. 3c, d, Extended Data Table 2).

Successful model realizations indicate that the concentration of regassed Xe retained in downwellings must have remained low (much lower than the mantle Xe concentration) until after about 2.5 Gyr ago (Fig. 4). Solutions correspond to a limited set of regassing histories, bounded by curves similar to C and D in Fig. 3: a near-zero $^{130}\text{Xe}_d$ that increases rapidly to a modest final $^{130}\text{Xe}_d$ in the last several hundred million years, and near-zero $^{130}\text{Xe}_d$ that increases to a sustained low magnitude over the past ~ 2 Gyr (Fig. 4a). On the basis of this analysis, the mantle shifted from net degassing to net regassing after about 2.5 Gyr ago (Fig. 4b, c). Accordingly, $^{130}\text{Xe}_d$ concentrations were low relative to the convecting mantle ^{130}Xe concentration until the Proterozoic at the earliest. Sustained low-magnitude $^{130}\text{Xe}_d$ (curve D in Fig. 3, light-blue curves in Fig. 4) yields an earlier shift to net regassing, whereas regassing histories with a late increase in $^{130}\text{Xe}_d$ yield a late shift to regassing and rapid recent change in mantle Xe isotopic composition (curve C in Fig. 3, dark-blue lines in Fig. 4, Extended Data Fig. 4). Constraints on Xe isotopic compositions in mantle-derived rocks over time would distinguish between viable regassing histories. We note that the results are robust among solutions derived using different initial mantle ^{130}Xe concentrations, mantle processing parameters and $^{130}\text{Xe}_d$ functional forms (Extended Data Figs. 5–7). In all successful model realizations, we find that the dominance of the modern atmospheric Xe isotopic signature in present-day mantle sources requires limited early regassing of ancient atmospheric Xe into the mantle (scenario (a) above).

Previous studies have shown that hydrous minerals in subducting lithologies carry Xe^{5,24,25} and that the abundance pattern of heavy noble gases in the mantle reflects incorporation of noble gases associated with hydrous minerals into the mantle^{5,26,27}. Because hydrous minerals carry Xe, H_2O cannot be regassed into the mantle without also regassing some Xe into the mantle. Limited early regassing of Xe therefore provides a constraint on early regassing of H_2O . The ratio of chemically bound H_2O to Xe varies among serpentinites, altered oceanic crust and sediments. If we take the distribution of $^{130}\text{Xe}_d$ values at 3 Gyr ago from successful model realizations and estimate the $\text{H}_2\text{O}/^{130}\text{Xe}$ ratio of serpentinite (a high $\text{H}_2\text{O}/\text{Xe}$ lithology) as 1.2×10^{13} (13 wt% H_2O and 3.7×10^8 atoms ^{130}Xe per gram⁵), then the median H_2O concentration in downwellings at 3 Gyr ago is ~ 0.61 p.p.m. H_2O and 95% of solutions have between ~ 0 and 62 p.p.m. H_2O . This range suggests very dry conditions compared to the estimated ~ 400 – $1,000$ p.p.m. H_2O concentration in present-day slabs subducting beyond depths of magma generation^{9,10,28}. If some Xe were regassed via materials with relatively

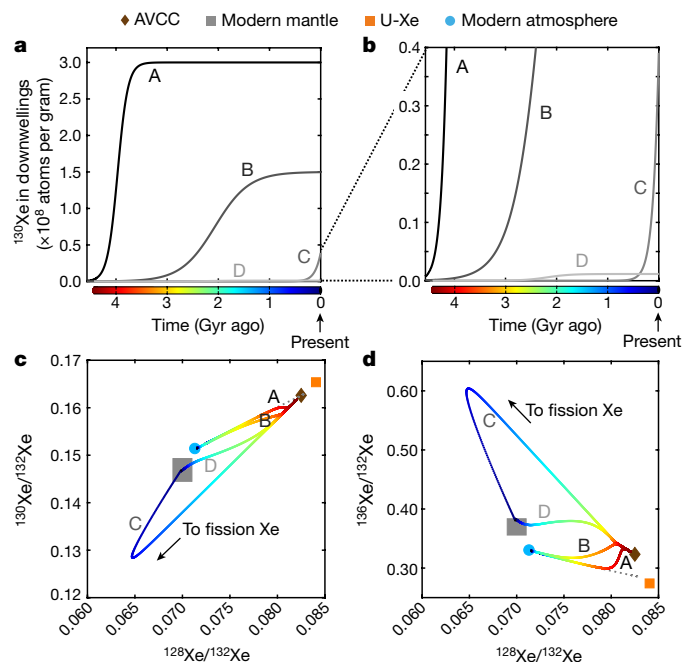
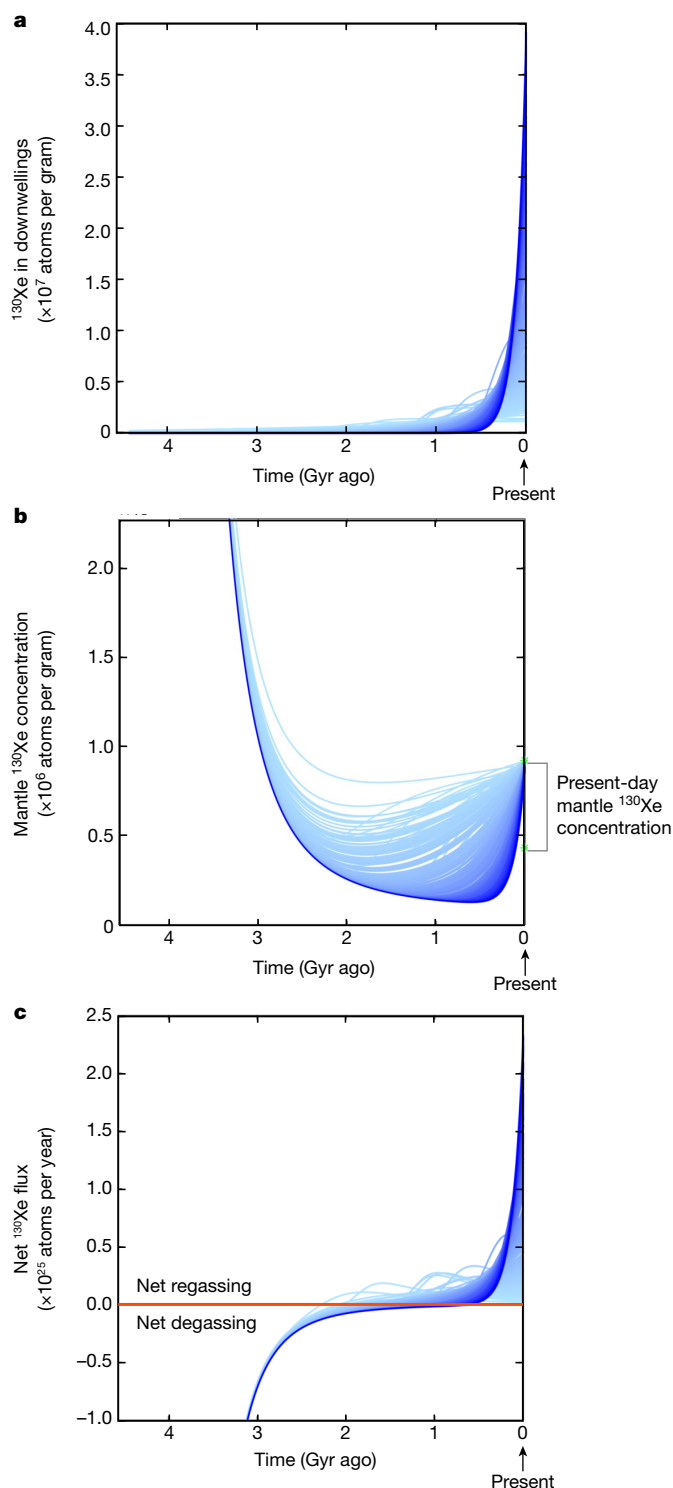


Fig. 3 | Mantle isotopic evolution curves. **a, b**, Four different regassing histories (A, B, C and D) are tested assuming an initial mantle ^{130}Xe concentration of 3.2×10^8 atoms per gram, a convecting mantle reservoir that is 90% of the mass of the whole mantle and 8 mantle-reservoir masses processed over Earth history. **c, d**, The corresponding mantle isotopic evolution curves (rainbow-coloured according to time) show how the mantle $^{128,130,132,136}\text{Xe}$ composition responds to degassing, regassing and fissionogenic production over time. The initial mantle composition is average carbonaceous chondrite (AVCC, brown diamond). The present-day mantle field is represented by a dark-grey box (Extended Data Table 2). Successful regassing histories yield a present-day mantle Xe isotopic composition and ^{130}Xe concentration within the field constrained by mantle Xe data. The atmosphere starts with a composition that is mass-fractionated by 39‰ per atomic mass unit relative to the modern atmosphere (similar to U-Xe^{17–19}, which is an estimate of the primordial atmospheric composition; orange square) and follows a Rayleigh mass-fractionation trend (dashed grey curve) to reach the modern atmospheric composition (light-blue circle) at 2.5 Gyr ago (Extended Data Fig. 1). Degassing drives mantle Xe isotope ratios towards Pu-fission and U-fission Xe compositions defined in Extended Data Table 2. Regassing drives mantle Xe isotope ratios towards the instantaneous atmospheric Xe composition. Strong regassing (curves A, B) puts too much atmospheric Xe in the mantle, so that the Xe isotopic composition of the mantle largely reflects the evolving atmospheric composition, despite fissionogenic production. The present-day mantle composition is achieved with regassing histories that have limited regassing until ~ 2.5 Gyr ago: either with negligible regassing through most of Earth history, increasing to modest regassing in the past few hundred million years (curve C), or with near-constant low-magnitude regassing over the past few billion years (curve D). Curves C and D represent the extremes of the successful regassing histories illustrated in Fig. 4a.

low $\text{H}_2\text{O}/\text{Xe}$, such as sediments, then the amount of water in Archaean downwellings would have been even lower. We note that as pressure and temperature increases, diffusion and advection of fluids released by hydrous mineral breakdown may alter $\text{H}_2\text{O}/\text{Xe}$ ratios in downwellings. If advection via hydrous-breakdown fluids removes Xe, then H_2O and Xe may remain coupled through subduction. Previous work indicates that dehydration of hydrous minerals may actually lower $\text{H}_2\text{O}/\text{Xe}$ ratios: dry olivine–enstatite residues formed by antigorite breakdown preserve Xe concentrations hundreds of times higher than that of the ambient upper mantle²⁷. If Xe is primarily carried in fluid inclusions, diffusion may explain this observation: H^+ is expected to diffuse out of inclusions more readily than Xe, potentially leading to preferential loss of H_2O relative to Xe at high pressures and temperatures^{5,27}. Low $\text{H}_2\text{O}/\text{Xe}$ ratios after the breakdown of hydrous minerals or desiccation



of fluid inclusions would further limit the concentration of H_2O that could have been regassed with Xe early in Earth history. Therefore, we suggest that on the basis of Xe isotopic constraints, downwellings before 2.5 Gyr ago were dry compared to modern-day subducting slabs.

We note that there is evidence for the initiation of subduction before the Proterozoic: eclogitic inclusions appear in diamonds 3.0 Gyr old and younger²⁹, and Hadean zircon thermobarometry suggests that magma production in convergent margin-like environments occurred as early as 4.2 Gyr ago³⁰. If plate tectonics and plate subduction were initiated before 2.5 Gyr ago, then early subducted material was either hydrated to a lesser extent at the surface than today, or volatiles were more efficiently expelled from Archaean slabs at shallow depths and returned to the surface. If high-temperature alteration of relatively

Fig. 4 | Xe regassing is limited early in Earth history. **a–c**, Time series showing regassing histories, $^{130}\text{Xe}_d$, for successful model realizations (**a**), the corresponding mantle ^{130}Xe concentrations over time (**b**) and the net flux between the convecting mantle and atmosphere over time (**c**). Successful model realizations indicate minimal ^{130}Xe in downwelling material until ~ 2.5 Gyr ago. Line colour in all panels reflects the present-day $^{130}\text{Xe}_d$ concentrations from **a**: darker blue indicates higher present-day $^{130}\text{Xe}_d$ concentrations. Successful realizations with relatively high present-day $^{130}\text{Xe}_d$ concentrations are associated with regassing onset times in the past few hundred million years (dark blue), whereas regassing histories with earlier onset times (1–2 Gyr ago) have very low present-day $^{130}\text{Xe}_d$ concentrations (light blue; see Extended Data Fig. 4). Successful model realizations indicate that the mantle shifted from a net degassing regime to a net regassing regime about 2.5 Gyr ago or later: $^{130}\text{Xe}_d$ concentrations were low relative to the convecting mantle ^{130}Xe concentration until the Proterozoic at the earliest (**b**, **c**; Extended Data Fig. 8). The successful model regassing histories with the lowest and highest present-day $^{130}\text{Xe}_d$ values (lightest and darkest blue curves, respectively) correspond to curves C and D in Fig. 3. Model results reflect an initial mantle ^{130}Xe concentration of 3.2×10^8 atoms per gram, a convecting mantle reservoir that is 90% of the mass of the whole mantle, 8 mantle-reservoir masses processed over Earth history and the continental crust growth model 1 (see Methods). Other parameter combinations and Xe fluxes are illustrated in Extended Data Figs. 5–7 and emphasize that regassing must have been limited early in Earth history.

thick Archaean crust in a Xe-rich early atmosphere promoted high initial ^{130}Xe concentrations in surface-altered materials, then the latter effect must have dominated in order to yield low $^{130}\text{Xe}_d$ concentrations. Independent of these physical factors, our results indicate that the convecting mantle experienced net degassing during the Archaean and transitioned to a net regassing regime after 2.5 Gyr ago.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0388-4>

Received: 23 August 2017; Accepted: 5 June 2018;

Published online 8 August 2018.

- Pujol, M., Marty, B. & Burgess, R. Chondritic-like xenon trapped in Archean rocks: a possible signature of the ancient atmosphere. *Earth Planet. Sci. Lett.* **308**, 298–306 (2011).
- Avice, G., Marty, B. & Burgess, R. The origin and degassing history of the Earth's atmosphere revealed by Archean xenon. *Nat. Commun.* **8**, 15455 (2017).
- Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry* (Cambridge University Press, Cambridge, 2002).
- Jackson, C. R., Parman, S. W., Kelley, S. P. & Cooper, R. F. Noble gas transport into the mantle facilitated by high solubility in amphibole. *Nat. Geosci.* **6**, 562–565 (2013).
- Kendrick, M. A. et al. Subduction zone fluxes of halogens and noble gases in seafloor and forearc serpentinites. *Earth Planet. Sci. Lett.* **365**, 86–96 (2013).
- Brown, M. Duality of thermal regimes is the distinctive characteristic of plate tectonics since the Neoproterozoic. *Geology* **34**, 961–964 (2006).
- Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
- Hacker, B. R. H_2O subduction beyond arcs. *Geochim. Geophys. Res.* **9**, Q03001 (2008).
- van Keken, P. E., Hacker, B. R., Syracuse, E. M. & Abers, G. A. Subduction factory: 4. Depth-dependent flux of H_2O from subducting slabs worldwide. *J. Geophys. Res.* **116**, B01401 (2011).
- Parai, R. & Mukhopadhyay, S. How large is the subducted water flux? New constraints on mantle regassing rates. *Earth Planet. Sci. Lett.* **317–318**, 396–406 (2012).
- Holland, G. & Ballentine, C. J. Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* **441**, 186–191 (2006).
- Mukhopadhyay, S. Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* **486**, 101–104 (2012).
- Parai, R., Mukhopadhyay, S. & Standish, J. J. Heterogeneous upper mantle Ne, Ar and Xe isotopic compositions and a possible Dupal noble gas signature recorded in basalts from the Southwest Indian Ridge. *Earth Planet. Sci. Lett.* **359–360**, 227–239 (2012).

14. Tucker, J. M., Mukhopadhyay, S. & Schilling, J. G. The heavy noble gas composition of the depleted MORB mantle (DMM) and its implications for the preservation of heterogeneities in the mantle. *Earth Planet. Sci. Lett.* **355–356**, 244–254 (2012).
15. Pető, M. K., Mukhopadhyay, S. & Kelley, K. A. Heterogeneities from the first 100 million years recorded in deep mantle noble gases from the Northern Lau Back-arc Basin. *Earth Planet. Sci. Lett.* **369–370**, 13–23 (2013).
16. Parai, R. & Mukhopadhyay, S. The evolution of MORB and plume mantle volatile budgets: constraints from fission Xe isotopes in Southwest Indian Ridge basalts. *Geochem. Geophys. Geosyst.* **16**, 719–735 (2015).
17. Pepin, R. O. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus* **92**, 2–79 (1991).
18. Pepin, R. O. On the isotopic composition of primordial xenon in terrestrial planet atmospheres. *Space Sci. Rev.* **92**, 371–395 (2000).
19. Pepin, R. O. & Porcelli, D. Origin of noble gases in the terrestrial planets. *Rev. Mineral. Geochem.* **47**, 191–246 (2002).
20. Pujol, M., Marty, B., Burnard, P. & Philippot, P. Xenon in Archean barite: weak decay of ^{130}Ba , mass-dependent isotopic fractionation and implication for barite formation. *Geochim. Cosmochim. Acta* **73**, 6834–6846 (2009).
21. Marty, B. The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).
22. Caracausi, A., Avice, G., Burnard, P. G., Füre, E. & Marty, B. Chondritic xenon in the Earth's mantle. *Nature* **533**, 82–85 (2016).
23. Caffee, M. et al. Primordial noble gases from Earth's mantle: identification of a primitive volatile component. *Science* **285**, 2115–2118 (1999).
24. Chavrit, D. et al. The contribution of hydrothermally altered ocean crust to the mantle halogen and noble gas cycles. *Geochim. Cosmochim. Acta* **183**, 106–124 (2016).
25. Matsuda, J. I. & Nagao, K. Noble gas abundances in a deep-sea sediment core from eastern equatorial Pacific. *Geochem. J.* **20**, 71–80 (1986).
26. Sumino, H. et al. Seawater-derived noble gases and halogens preserved in exhumed mantle wedge peridotite. *Earth Planet. Sci. Lett.* **294**, 163–172 (2010).
27. Kendrick, M. A., Scambelluri, M., Honda, M. & Phillips, D. High abundances of noble gas and chlorine delivered to the mantle by serpentinite subduction. *Nat. Geosci.* **4**, 807–812 (2011).
28. Korenaga, J., Planavsky, N. J. & Evans, D. A. Global water cycle and the coevolution of the Earth's interior and surface environment. *Phil. Trans. R. Soc. A* **375**, 20150393 (2017).
29. Shirey, S. B. & Richardson, S. H. Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* **333**, 434–436 (2011).
30. Hopkins, M., Harrison, T. M. & Manning, C. E. Low heat flow inferred from >4 Gyr zircons suggests Hadean plate boundary interactions. *Nature* **456**, 493–496 (2008).

Acknowledgements The project was funded by NSF grant EAR-1250419 to S.M. We thank D. Fike, C. Jackson, M. Krawczynski and S. Turner for discussions that improved the manuscript.

Reviewer information Nature thanks M. Kendrick and D. Porcelli for their contribution to the peer review of this work.

Author contributions S.M. and R.P. developed the conceptual model and the ideas presented in the manuscript. R.P. wrote the Matlab scripts for the numerical modelling. R.P. and S.M. analysed the results and R.P. wrote the manuscript with input from S.M.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0388-4>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to R.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Linear least-squares fits of mantle source Xe using Archaean atmosphere.

Mantle source $^{130,131,132,134,136}\text{Xe}$ compositions are modelled as four-component mixtures of initial mantle Xe, recycled atmospheric Xe, Pu-fission Xe and U-fission Xe using the method outlined in ref. ¹⁶. We test two end-member scenarios: the regassed atmospheric Xe is either entirely modern or entirely ancient in its isotopic composition. We estimate the $^{130,131,132,134,136}\text{Xe}$ isotopic composition of Archaean atmosphere based on constraints from fluid inclusions in ancient rocks². We take the present-day atmospheric Xe composition and apply Rayleigh mass-dependent fractionation of 20‰ per atomic mass unit (AMU). The resulting composition is consistent with the atmospheric composition determined for Barberton samples after correction for fission production after closure of the fluid inclusions². The 20‰ AMU⁻¹ fractionated model Archaean atmospheric Xe composition used for comparing the goodness of fit for the convecting mantle source composition is given in Extended Data Fig. 2e.

A sum of squared residuals of zero indicates that the mantle source composition can be fitted perfectly by mixing the four end-member components. Sums of squared residuals greater than zero indicate the sigma-normalized error in the best fit to the mantle source composition. If modern atmospheric Xe is taken as the recycled atmospheric Xe component, sums of squared residuals are zero or near-zero. If 20‰ AMU⁻¹ mass-fractionated ancient atmosphere is the recycled component, then the sums of squared residuals are much higher than zero, indicating that mantle source compositions cannot be explained by recycling of only ancient atmospheric Xe (Extended Data Fig. 2).

Model initialization. The concentration and isotopic composition of Xe in the mantle is tracked over time in a numerical forward model of mantle degassing, regassing and fissionogenic production. The initial concentration of ^{130}Xe represents the primordial Xe component delivered and retained throughout accretion. Initial mantle ^{130}Xe concentrations corresponding to Xe contributions from carbonaceous chondrites in a late veneer of 0.1%, 0.5% and 1% of the Earth's mass are tested. Figures 3, 4 correspond to model runs with a late veneer equivalent to 1% of the Earth's mass with a carbonaceous chondrite ^{130}Xe concentration (based on the average of Murchison and Orgueil²¹). The initial mantle Xe isotopic composition is taken to be that of average carbonaceous chondrite^{18,22}. Results for a late veneer corresponding to 0.1% and 0.5% of the Earth's mass are illustrated in Extended Data Figs. 6, 7. The initial atmospheric Xe isotopic composition is mass-fractionated with respect to modern atmosphere by 39‰ AMU⁻¹.

Xe concentrations in downwellings over time. We use a Monte Carlo method to explore two functional forms for the concentration of ^{130}Xe in downwelling material over time. The first is a sigmoidal function based on the generalized logistic function:

$$^{130}\text{Xe}_d(t) = \frac{^{130}\text{Xe}_d^k}{1 + e^{-\alpha(t-\beta)}} \quad (1)$$

where $^{130}\text{Xe}_d^k$ is the carrying-capacity ^{130}Xe concentration in downwellings, α is the growth rate and β is the sigmoid inflection point. Figure 2 illustrates an example array of sigmoidal $^{130}\text{Xe}_d$ functions tested with the model. Each sigmoid is sampled over a limited time interval and thus yields only a portion of the sigmoid shape. Therefore, the initial $^{130}\text{Xe}_d$ values may be greater than 0 and the present-day $^{130}\text{Xe}_d$ values are lower than or equal to the carrying capacity, $^{130}\text{Xe}_d^k$.

The second functional form is an exponential form:

$$^{130}\text{Xe}_d(t) = ^{130}\text{Xe}_d^{\text{final}} e^{\tau(t-T)} \quad (2)$$

where T is the age of the Earth (4.568 Gyr) and the two free parameters are $^{130}\text{Xe}_d^{\text{final}}$ and a time constant, τ . Time constant values between 10^{-11} Gyr⁻¹ and 5×10^{-8} Gyr⁻¹ are explored. Extended Data Fig. 3 illustrates a coarse array of exponential $^{130}\text{Xe}_d$ functions tested with the model. For both functional forms, we use a Monte Carlo numerical method to achieve good coverage of the free parameter space.

The present-day $^{130}\text{Xe}_d$ concentrations tested range from zero to 5×10^8 atoms per gram. We reiterate that regassing refers to the influx of volatiles that are transported beyond depths of magma generation and mixed into the convecting mantle. We place a broad upper limit on the present-day amount of ^{130}Xe in downwelling material to constrain the collection of $^{130}\text{Xe}_d$ time series tested with our model. Using ^{130}Xe concentration data from serpentinite, altered oceanic crust and sediments^{5,24,25,27,31–35}, we compute an upper limit on present-day regassing of Xe from constraints on regassing of water beyond depths of magma generation. We assume that serpentinite has 13 wt% H₂O, altered oceanic crust has 1.2 wt% H₂O³⁶ and average subducting sediment has 7.3 wt% H₂O³⁷. On the basis of mantle outgassing estimates and sea level constraints, ref. ¹⁰ determined an upper-limit H₂O flux of 7.0×10^{14} g yr⁻¹ beyond depths of magma generation, corresponding to a sea level decrease of 360 m over the Phanerozoic. If this full upper limit is carried in

sediments with 7.3 wt% H₂O, then this flux corresponds to 9.6×10^{15} g yr⁻¹ of sediment. Using the maximum sedimentary ^{130}Xe concentration of 3.2×10^{10} atoms per gram, we compute an upper-limit present-day ^{130}Xe flux beyond depths of magma generation of 3.0×10^{26} atoms ^{130}Xe per year. This flux is then distributed over the present-day mass of downwelling per year (6.1×10^{17} g yr⁻¹) to yield a maximum present-day bulk downwelling ^{130}Xe concentration of 5.0×10^8 atoms per gram. Because the maximum present-day $^{130}\text{Xe}_d$ in successful model realizations is much lower ($< 4 \times 10^7$ atoms per gram for sigmoidal $^{130}\text{Xe}_d$ and $< 2 \times 10^8$ atoms per gram for exponential $^{130}\text{Xe}_d$), we note that this estimated upper limit only helps to define our $^{130}\text{Xe}_d$ parameter search windows.

Time evolution. Degassing occurs during partial melting of upwelling convecting mantle, and regassing occurs via the corresponding downwellings (presently at mid-ocean ridges and subduction zones, respectively). We test an exponentially decreasing mantle processing rate. The mantle processing rate is tied to the present-day ridge processing rate, Q_p (6.1×10^{17} g yr⁻¹, assuming 10% partial melting to produce 21 km³ yr⁻¹ of crust at mid-ocean ridges, with a crustal density of 2,900 kg m⁻³)³⁸:

$$Q(t) = Q_p e^{\eta(T-t)} \quad (3)$$

where $T = 4.568$ Gyr. To test different mantle processing histories with our model, we test discrete values of η corresponding to whole numbers of mantle reservoir masses processed over Earth history (for example, $\eta = 8.1 \times 10^{-10}$ corresponds to $N_{\text{res}} = 8$ with $M_{\text{res}} = 90\%$). We test values from $\eta = 1.6 \times 10^{-10}$ to $\eta = 9.9 \times 10^{-10}$ (2–15 mantle reservoir masses ranging from 50% to 90% of the whole mantle). A linearly decreasing mantle processing rate yields similar final results. Thus, our broad conclusions are not sensitive to the functional form of the mantle processing-rate history.

For each time step, the mass of mantle (dM) processed between time t_{last} and t_{now} is given by:

$$dM = \left(\frac{Q_p}{\eta} \right) (e^{\eta(T-t_{\text{last}})} - e^{\eta(T-t_{\text{now}})}) \quad (4)$$

The normalized mass processed per time step is dM/M_{res} , where M_{res} is the mass of the convecting mantle. Results shown in the main-text figures are obtained by taking M_{res} to be 90% of the mass of the mantle, or 3.6×10^{27} g. Sensitivity to the mass of the convecting mantle is shown in Extended Data Figs. 5–7.

Over a given time step, the net change in mantle ^{130}Xe concentration corresponds to the balance between the ^{130}Xe lost by degassing of dM by partial melting and the ^{130}Xe gained by regassing via a corresponding mass of downwelling mantle with Xe concentration $^{130}\text{Xe}_d$ (Fig. 1, Extended Data Fig. 8). The concentration of ^{130}Xe in the mantle thus evolves according to:

$$^{130}\text{Xe}_{\text{now}}^m = ^{130}\text{Xe}_{\text{last}}^m + \left(\frac{dM}{M_{\text{res}}} \right) (^{130}\text{Xe}_{\text{last}}^d - ^{130}\text{Xe}_{\text{last}}^m) \quad (5)$$

where the superscript m denotes the concentration in the convecting mantle and the superscript d denotes the concentration in the downwelling material (Figs. 1, 2). The mantle ^{128}Xe concentration evolves similarly and is coupled to ^{130}Xe via the instantaneous $^{128}\text{Xe}/^{130}\text{Xe}$ ratios of the mantle and atmosphere (Extended Data Fig. 1):

$$^{128}\text{Xe}_{\text{now}}^m = ^{128}\text{Xe}_{\text{last}}^m + \left(\frac{dM}{M_{\text{res}}} \right) \left(^{130}\text{Xe}_{\text{last}}^d \left(\frac{^{128}\text{Xe}}{^{130}\text{Xe}} \right)_{\text{last}}^{\text{atm}} - ^{128}\text{Xe}_{\text{last}}^m \right) \quad (6)$$

Expressions for $^{131,132,134,136}\text{Xe}$ must additionally account for in situ production by spontaneous fission of ^{244}Pu and ^{238}U in the mantle. Bulk upper-mantle abundances of incompatible lithophile elements changed over Earth history in association with the growth of the continental crust. To model the depletion of the convecting mantle over time (extraction via partial melting, offset by recycling via downwellings), net U and Pu loss from the mantle at each time step tracks continental crustal growth over time (Extended Data Fig. 3).

We use three continental crust growth models (CCs) to test a range of growth rates similar to those proposed in the literature^{39–41} (Extended Data Fig. 3). Two sigmoidal growth curves are adopted: one with relatively rapid growth (CC = 1) and one with more protracted crustal growth (CC = 2). The third growth model (CC = 3) builds the continental crust at a constant rate (linear growth), beginning 300 Myr after the formation of the Solar System. We assume that the extraction of U and Pu is directly proportional to the extraction of continental crust from the convecting mantle reservoir over time. For each combination of M_{res} and CC, we solve for the unique scaling factor X that yields a present-day U concentration⁴² of 1.3 p.p.m. in the continental crust reservoir (mass of 2.2×10^{25} g)^{40,42}, given a total

bulk (depleted convecting mantle plus continental crust) U concentration equal to the bulk silicate Earth value of 21 p.p.b. $U^{43,44}$. We assume that Pu and U are not fractionated from one another by crustal extraction; thus, Pu extraction tracks U extraction using the atomic ratio of the two radioactive species at any given time. On the basis of the continental crust model, which gives the mass of the continental crust over time (normalized to 1 at the present; Extended Data Fig. 3), we derive the expression for the reduction in ^{238}U concentration in the convecting mantle over a given time step:

$$dCC = CC_{\text{now}} - CC_{\text{last}} \quad (7)$$

$$dU_{CC} = X \times dCC \quad (8)$$

where X is the scaling factor for ^{238}U extraction for a given M_{res} and CC . The expressions for the concentration of U and Pu in the convecting mantle then reflect decay and the decrease in the ^{238}U concentration over a given time step due to crustal extraction, dU_{CC} (in atoms per gram):

$$dt = t_{\text{now}} - t_{\text{last}} \quad (9)$$

$$^{238}U_{\text{now}} = ^{238}U_{\text{last}} e^{-\lambda_{238} dt} - dU_{CC} \quad (10)$$

$$^{244}Pu_{\text{now}} = ^{244}Pu_{\text{last}} e^{-\lambda_{244} dt} - \left(\frac{^{244}Pu}{^{238}U} \right)_{\text{last}} dU_{CC} \quad (11)$$

where $\lambda_{244} = 8.6643 \times 10^{-9} \text{ yr}^{-1}$ and $\lambda_{238} = 1.5514 \times 10^{-10} \text{ yr}^{-1}$. The model is very weakly sensitive to the continental crust growth model (Extended Data Fig. 7).

The time evolution of a fissiogenic xenon isotope $^{\psi}\text{Xe}$ can thus be broken down into four equations, reflecting the primordial mantle (p.m.) budget, regassed mantle (r.m.) budget, Pu-fissiogenic budget and U-fissiogenic budget, respectively:

$$^{\psi}\text{Xe}_{\text{now}}^{\text{p.m.}} = ^{\psi}\text{Xe}_{\text{last}}^{\text{p.m.}} \left(1 - \frac{dM}{M_{\text{res}}} \right) \quad (12)$$

$$^{\psi}\text{Xe}_{\text{now}}^{\text{r.m.}} = ^{\psi}\text{Xe}_{\text{last}}^{\text{r.m.}} + \left(\frac{dM}{M_{\text{res}}} \right) \left[\left(^{130}\text{Xe}_{\text{last}}^{\text{d}} \left(\frac{^{\psi}\text{Xe}}{^{130}\text{Xe}} \right)_{\text{last}}^{\text{atm}} \right) - ^{\psi}\text{Xe}_{\text{last}}^{\text{r.m.}} \right] \quad (13)$$

$$^{\psi}\text{Xe}_{\text{now}}^{\text{Pu}} = [^{\psi}\text{Xe}_{\text{last}}^{\text{Pu}} + ^{244}\text{Pu}_{\text{last}} (1 - e^{-\lambda_{244} dt}) Y_{\psi}^{\text{Pu}}] \left(1 - \frac{dM}{M_{\text{res}}} \right) \quad (14)$$

$$^{\psi}\text{Xe}_{\text{now}}^{\text{U}} = [^{\psi}\text{Xe}_{\text{last}}^{\text{U}} + ^{238}\text{U}_{\text{last}} (1 - e^{-\lambda_{238} dt}) Y_{\psi}^{\text{U}}] \left(1 - \frac{dM}{M_{\text{res}}} \right) \quad (15)$$

$$^{\psi}\text{Xe}_{\text{now}}^{\text{total}} = ^{\psi}\text{Xe}_{\text{now}}^{\text{p.m.}} + ^{\psi}\text{Xe}_{\text{now}}^{\text{r.m.}} + ^{\psi}\text{Xe}_{\text{now}}^{\text{Pu}} + ^{\psi}\text{Xe}_{\text{now}}^{\text{U}} \quad (16)$$

where $\psi = \{131, 132, 134, 136\}$.

The fission yield of ^{136}Xe from ^{244}Pu (Y_{136}^{Pu}) is taken to be 7×10^{-5} , and the fission yield of ^{136}Xe from ^{238}U (Y_{136}^{U}) is taken to be 3.43×10^{-8} (ref. 45). Yields for the other fission isotopes of Xe, calculated on the basis of fissiogenic Xe spectra, are given in Extended Data Table 2. The initial $^{244}\text{Pu}/^{238}\text{U}$ ratio is taken to be 0.0068 (ref. 46).

The diverse half-lives of tracked radioactive species (^{129}I , ^{244}Pu and ^{238}U) necessitate a tailored time step scheme: the time step must be fine enough to accurately capture the radioactive decay of short-lived ^{244}Pu early in Earth history, but after ~ 1 Gyr, a very fine time step is not required for accuracy and imposes a high computational cost. We carried out convergence tests to determine the optimal time resolution scheme that accurately captures the decay of ^{244}Pu and ^{238}U to at least three significant figures. On the basis of this analysis, the time step is 0.1 Myr from the time when accretion is completed until 200 Myr after the formation of the Solar System, 1 Myr from 4.368 Gyr ago until 3.3 Gyr ago, and 5 Myr through the rest of Earth history to the present.

Model success criterion 1: present-day convecting mantle ^{130}Xe concentration. The present-day concentration of ^{130}Xe in the convecting mantle is estimated using the ^3He mantle outgassing flux, the model present-day mantle processing rate and the $^{130}\text{Xe}/^3\text{He}$ ratio of the mid-ocean ridge basalt mantle source. Estimates of the mantle outgassing flux vary, and a range of 400–850 moles of ^3He per year covers recent estimates^{47–49}. Our model present-day mantle processing rate is based on $21 \text{ km}^3 \text{ yr}^{-1}$ of oceanic crust production³⁸, assuming an average crustal density of

2.9 g cm^{-3} and 10% partial melting to produce oceanic crust on average⁵⁰. The convecting mantle $^{130}\text{Xe}/^3\text{He}$ ratio is estimated to be 1.1×10^{-3} based on a robust fit of $^{129}\text{Xe}/^{130}\text{Xe}$ versus $^3\text{He}/^{130}\text{Xe}$ data from the popping rock 2IID43 (ref. 51) and taking the mantle source $^{129}\text{Xe}/^{130}\text{Xe}$ ratio¹² to be 7.8. Using these values, we determine a convecting mantle ^{130}Xe concentration range of 4.3×10^5 to 9.2×10^5 atoms ^{130}Xe per gram. This estimate is about two times lower than another recent estimate²¹. However, Marty²¹ uses a higher ridge ^3He outgassing flux of $1,000 \pm 250$ moles per year⁵² to constrain the absolute concentration of ^{130}Xe in the upper mantle. Halliday⁵³ also provides an estimate of mantle ^{130}Xe concentration, but this is an estimate of the bulk mantle based partially on data from plume-derived samples. Model realizations that produce present-day convecting mantle ^{130}Xe concentrations between 4.3×10^5 and 9.2×10^5 atoms per gram are considered successful.

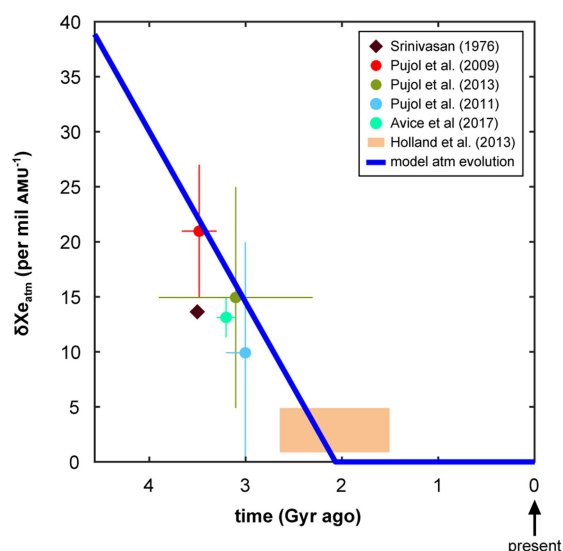
Model success criterion 2: present-day convecting mantle Xe isotopic composition. The primordial mantle Xe isotopic composition is taken to be chondritic²². In successful model realizations, the mantle Xe isotopic composition evolves from average carbonaceous chondrite^{17,18} to the present-day mantle range via the addition of Xe derived from Pu fission and U fission, regassing of atmospheric Xe that evolves as a function of time (Extended Data Fig. 1) and degassing. Constraints on the convecting mantle $^{128}\text{Xe}/^{130}\text{Xe}$ ratio are based on measurements in well gases and mantle-derived basalts. $^{128}\text{Xe}/^{130}\text{Xe}$ ratios up to ~ 0.478 are measured in well gases¹¹. Model realizations that produce present-day mantle $^{128}\text{Xe}/^{130}\text{Xe}$ ratios between 0.475 and 0.478 are considered successful. Constraints on the ^{132}Xe -normalized Xe isotope ratios are based on measurements of mantle-derived basalts^{14,16} and well gases^{11,23}. Extended Data Table 2 gives the present-day range of mantle Xe isotopic compositions used to determine successful model realizations.

Code availability. A Matlab code for modelling mantle Xe isotopic evolution is available from the authors upon reasonable request.

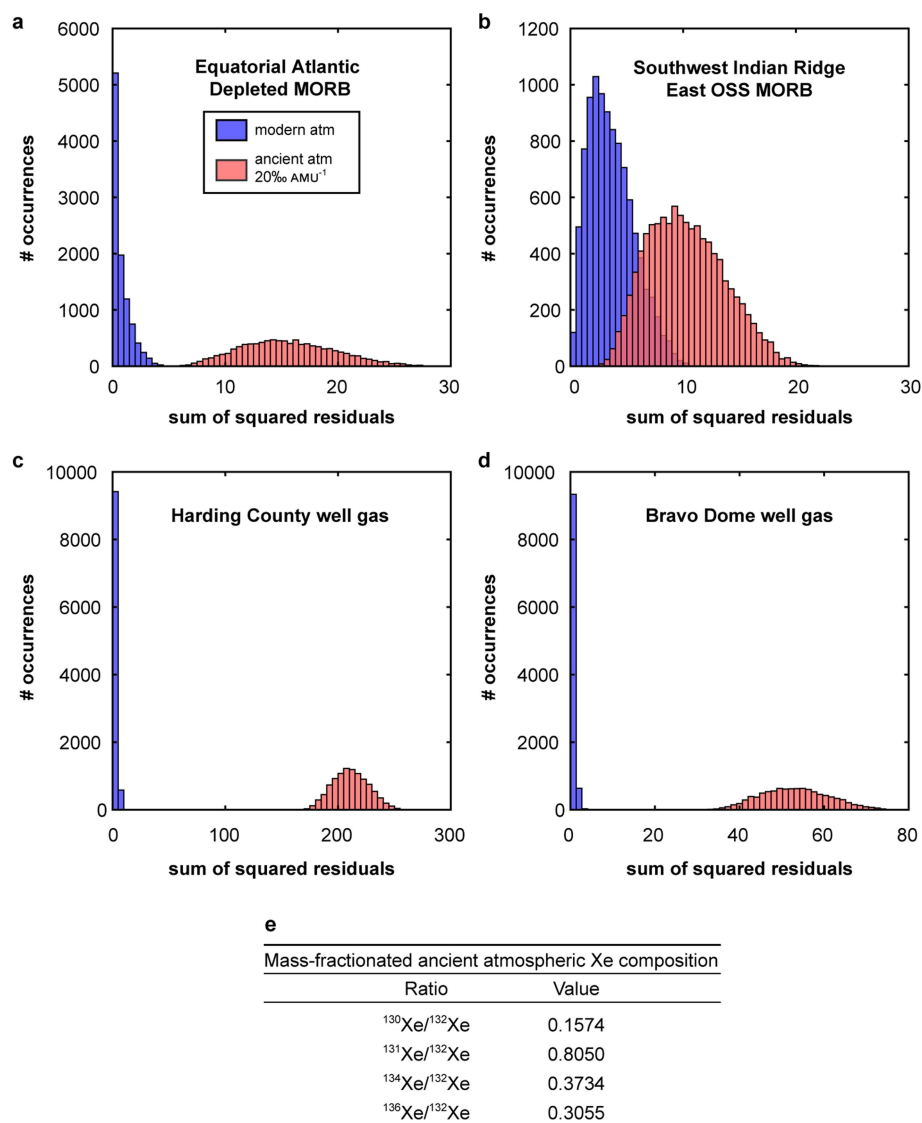
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

- Kumagai, H., Dick, H. J. & Kaneoka, I. Noble gas signatures of abyssal gabbros and peridotites at an Indian Ocean core complex. *Geochem. Geophys. Geosyst.* **4**, 9017 (2003).
- Matsuda, J. I. & Matsubara, K. Noble gases in silica and their implication for the terrestrial “missing” Xe. *Geophys. Res. Lett.* **16**, 81–84 (1989).
- Matsuda, J. I. & von Herzen, R. Thermal conductivity variation in a deep-sea sediment core and its relation to H_2O , Ca and Si content. *Deep-Sea Res. A* **33**, 165–175 (1986).
- Podosek, F., Honda, M. & Ozima, M. Sedimentary noble gases. *Geochim. Cosmochim. Acta* **44**, 1875–1884 (1980).
- Staudacher, T. & Allègre, C. J. Recycling of oceanic crust and sediments: the noble gas subduction barrier. *Earth Planet. Sci. Lett.* **89**, 173–183 (1988).
- Jarrard, R. D. Subduction fluxes of water, carbon dioxide, chlorine, and potassium. *Geochem. Geophys. Geosyst.* **4**, 8905 (2003).
- Plank, T. & Langmuir, C. H. The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chem. Geol.* **145**, 325–394 (1998).
- Crisp, J. A. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20**, 177–211 (1984).
- Dhuime, B., Hawkesworth, C. J., Cawood, P. A. & Storey, C. D. A change in the geodynamics of continental growth 3 billion years ago. *Science* **335**, 1334–1336 (2012).
- McLennan, S. M. & Taylor, S. R. Geochemical constraints on the growth of the continental crust. *J. Geol.* **90**, 347–361 (1982).
- Pujol, M., Marty, B., Burgess, R., Turner, G. & Philippot, P. Argon isotopic composition of Archean atmosphere probes early Earth geodynamics. *Nature* **498**, 87–90 (2013).
- Rudnick, R. & Gao, S. *Treatise on Geochemistry* Vol. 3 (eds Holland, H. D. & Turekian, K. K.) 1–64 (Elsevier, 2003).
- McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Palme, H. & O'Neill, H. S. C. in *Treatise on Geochemistry* Vol. 2 (eds Holland, H. D. & Turekian, K. K.) 1–38 (Elsevier, Amsterdam, 2004).
- Tolstikhin, I., Marty, B., Porcelli, D. & Hofmann, A. Evolution of volatile species in the Earth's mantle: a view from xenology. *Geochim. Cosmochim. Acta* **136**, 229–246 (2014).
- Hudson, G. B., Kennedy, B. M., Podosek, F. A. & Hohenberg, C. M. In *Proc. 19th Lunar and Planetary Science Conference* 547–557 (Lunar and Planetary Institute, 1989).
- Bianchi, D. et al. Low helium flux from the mantle inferred from simulations of oceanic helium isotope data. *Earth Planet. Sci. Lett.* **297**, 379–386 (2010).
- Holzer, M. et al. Objective estimates of mantle ^3He in the ocean and implications for constraining the deep ocean circulation. *Earth Planet. Sci. Lett.* **458**, 305–314 (2017).
- Schlitzer, R. Quantifying He fluxes from the mantle using multi-tracer data assimilation. *Phil. Trans. R. Soc. A* **374**, 20150288 (2016).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).

51. Moreira, M., Kunz, J. & Allegre, C. Rare gas systematics in Popping Rock: isotopic and elemental compositions in the upper mantle. *Science* **279**, 1178–1181 (1998).
52. Craig, H., Clarke, W. & Beg, M. Excess ^3He in deep water on the East Pacific Rise. *Earth Planet. Sci. Lett.* **26**, 125–132 (1975).
53. Halliday, A. N. The origins of volatiles in the terrestrial planets. *Geochim. Cosmochim. Acta* **105**, 146–171 (2013).
54. Holland, G. et al. Deep fracture fluids isolated in the crust since the Precambrian era. *Nature* **497**, 357–360 (2013).
55. Srinivasan, B. Barites: anomalous xenon from spallation and neutron-induced reactions. *Earth Planet. Sci. Lett.* **31**, 129–141 (1976).
56. Alexander, E. C. Jr, Lewis, R. S., Reynolds, J. H. & Michel, M. C. Plutonium-244: confirmation as an extinct radioactivity. *Science* **172**, 837–840 (1971).
57. Lewis, R. S. Rare-gases in separated whitlockite from St. Severin chondrites: xenon and krypton from fission extinct Pu-244. *Geochim. Cosmochim. Acta* **39**, 417–432 (1975).
58. Wetherill, G. W. Spontaneous fission yields from uranium and thorium. *Phys. Rev.* **92**, 907–912 (1953).
59. Hebeda, E. H., Schultz, L. & Freundel, M. Radiogenic, fissiogenic and nucleogenic noble gases in zircons. *Earth Planet. Sci. Lett.* **85**, 79–90 (1987).
60. Eikenberg, J., Signer, P. & Wieler, R. U-Xe, U-Kr, AND U-Pb systematics for dating uranium minerals and investigations of the production of nucleogenic neon and argon. *Geochim. Cosmochim. Acta* **57**, 1053–1069 (1993).
61. Ragettli, R. A., Hebeda, E. H., Signer, P. & Wieler, R. Uranium xenon chronology: precise determination of $\lambda_{\text{sf}}^{136}\text{V}_{\text{sf}}$ for spontaneous fission of ^{238}U . *Earth Planet. Sci. Lett.* **128**, 653–670 (1994).

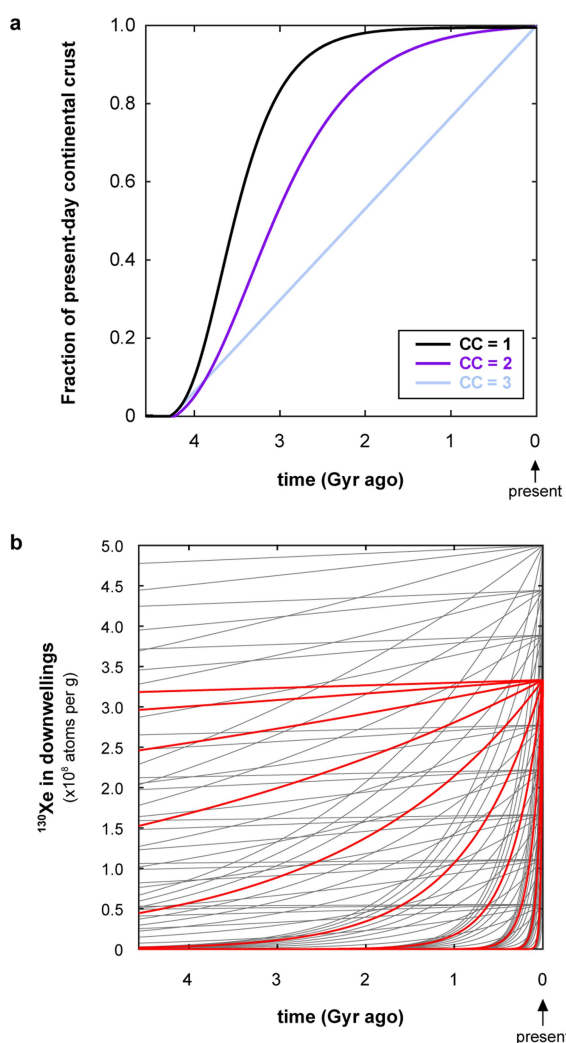


Extended Data Fig. 1 | Atmospheric Xe mass fractionation relative to the modern composition over time. Figure adapted from ref. ². Xe measured in Archaean barites, fluid inclusions in quartz from Archaean cherts and deep crustal fluids of various age are shown with associated 2σ uncertainties^{1,2,20,41,54,55}. The blue line shows the model atmospheric Xe mass fractionation over time. We assume that the initial Xe isotopic composition of the atmosphere is Rayleigh-mass-fractionated by $\sim 39\text{‰ AMU}^{-1}$ relative to the modern atmosphere and that the degree of mass fractionation decreases linearly until 2 Gyr ago (Ga), when the atmosphere reaches its present composition.

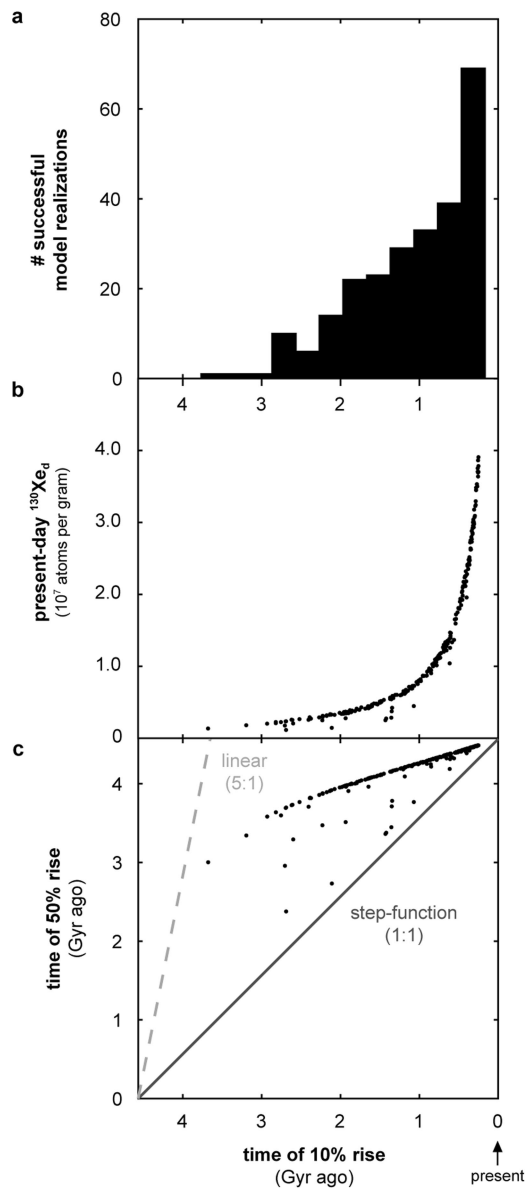


Extended Data Fig. 2 | Sum of squared residuals from least-squares fitting of mantle source compositions using either modern or ancient atmospheric Xe. Mantle source $^{130,131,132,134,136}\text{Xe}$ compositions are modelled as four-component mixtures of initial mantle Xe, recycled atmospheric Xe, and Xe from the fission of Pu and U. A sum of squared residuals of zero indicates that the mantle source composition can be fitted perfectly by mixing the four end-member components. Sums of squared residuals greater than zero indicate the sigma-normalized error in the best fit compared to the mantle source composition. Using modern atmospheric Xe as the regassed atmospheric Xe component, sums of squared residuals are zero or near-zero. Using ancient atmosphere, sums of squared residuals are much higher than zero, indicating that mantle

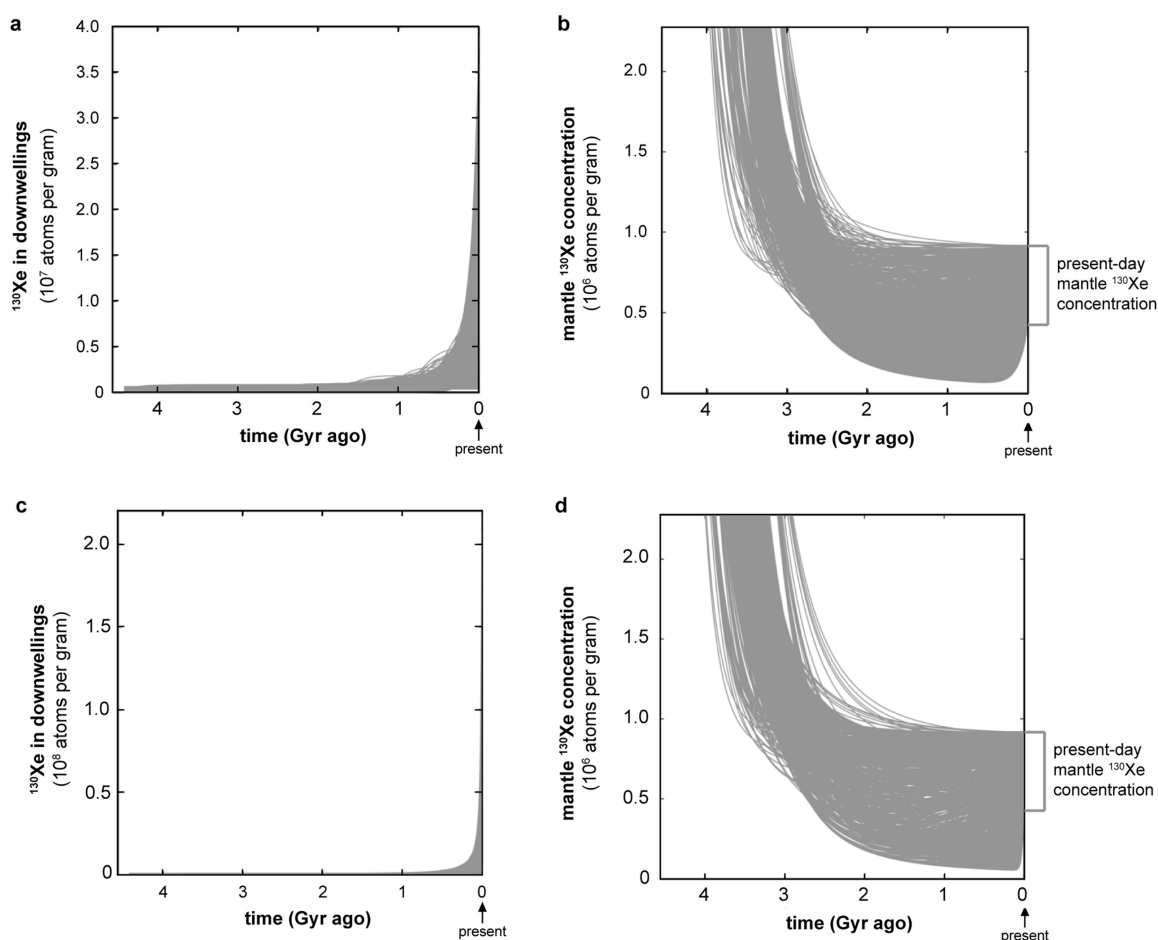
source compositions cannot be explained by regassing of only ancient atmospheric Xe. The ancient atmospheric Xe composition used here corresponds to 20‰ AMU^{-1} Rayleigh fractionation applied to the modern atmospheric composition and agrees with fission-corrected ancient atmosphere derived from fluid inclusions in Archaean rocks². **a–e**, Mantle source compositions for Equatorial Atlantic depleted mid-ocean ridge basalt (MORB)¹⁴ (**a**), Southwest Indian Ridge Eastern Orthogonal Supersegment MORB¹⁶ (**b**), Harding County well gas²³ (**c**) and Bravo Dome well gas¹¹ (**d**) are fitted using the Monte Carlo method ($n = 10,000$) described in ref. ¹⁶, with average carbonaceous chondrite as the initial mantle composition²², and either modern or 20‰ AMU^{-1} fractionated atmosphere (**e**) as the recycled component.



Extended Data Fig. 3 | Continental crust growth models and exponential $^{130}\text{Xe}_d$ time series examples. **a**, ^{238}U (and a small amount of ^{244}Pu) extraction from the mantle by partial melting is offset by recycling of sediments at subduction zones at each time step. We model net extraction of U and any extant Pu from the mantle as directly tracking continental crust growth over time (Methods). Three CCs are adopted: two sigmoidal curves that approximate literature continental crust growth curves ('CC = 1' and 'CC = 2') and one linear growth curve ('CC = 3')^{39–41}. **b**, Example of exponential $^{130}\text{Xe}_d$ time series tested with our forward model of mantle Xe evolution. Two parameters describe the exponential function (Methods): $^{130}\text{Xe}_d^{\text{final}}$, the final ^{130}Xe concentration in downwellings, and τ , the exponential time constant. Grey lines represent a collection of exponentials with discrete variation in $^{130}\text{Xe}_d^{\text{final}}$ and τ . A subset with a constant $^{130}\text{Xe}_d^{\text{final}}$ and varying τ is highlighted in red. The time constant τ is varied from $10^{-11} \text{ Gyr}^{-1}$ to $5 \times 10^{-8} \text{ Gyr}^{-1}$, with small τ corresponding to slow growth. Examples for nine different $^{130}\text{Xe}_d^{\text{final}}$ values are shown, with an upper bound of 5×10^8 atoms ^{130}Xe per gram (Methods).

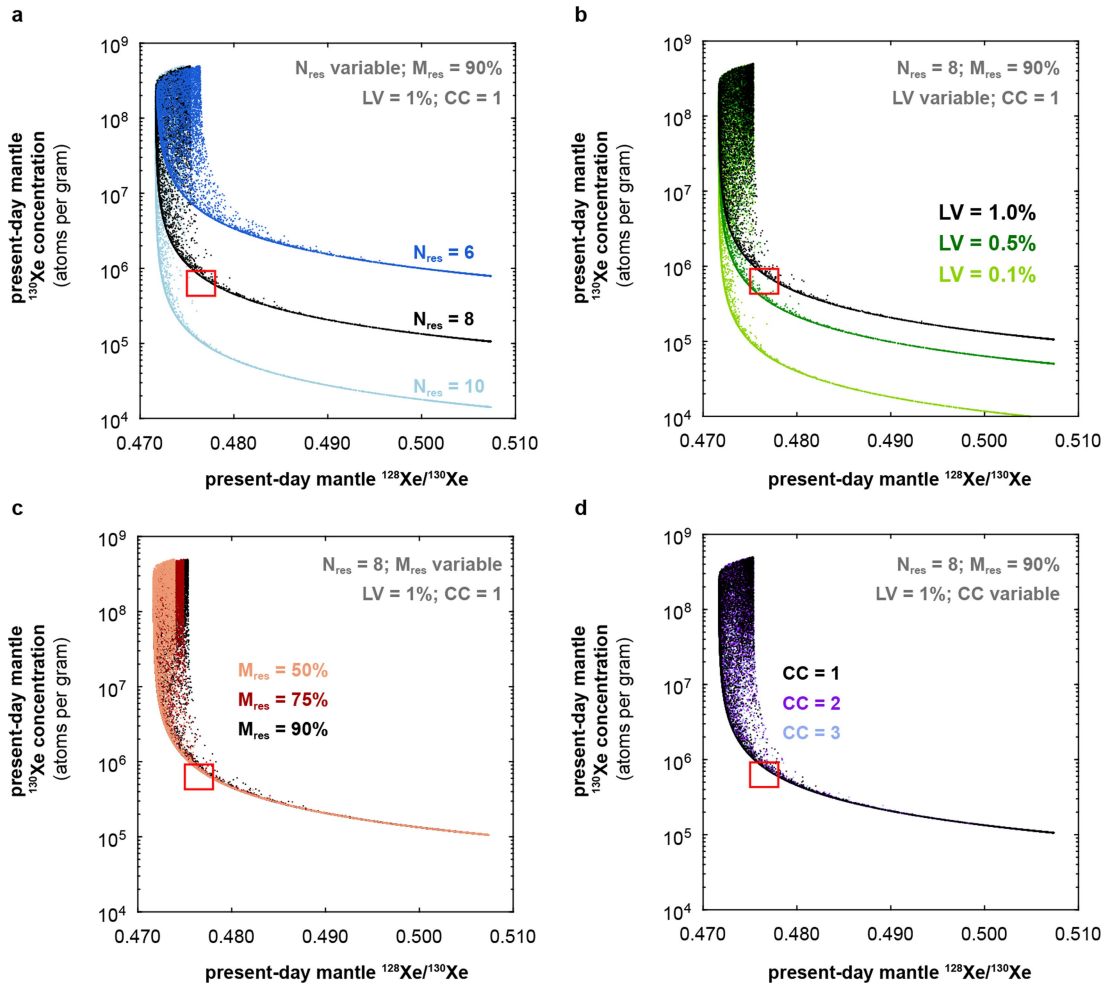


Extended Data Fig. 4 | Characterization of successful regassing histories. Diverse regassing history shapes are generated by sampling a limited time interval for a variety of growth rates and inflection times (Fig. 2). To provide a common point of comparison for the evolving conditions within downwellings, we sort results by the time when $^{130}\text{Xe}_d$ has increased by 10% between its initial and final values (time of 10% rise). **a**, Times of 10% rise for successful regassing histories. Most successful model realizations have a time of 10% rise later than 2.5 Gyr ago. **b**, Model realizations with high present-day $^{130}\text{Xe}_d$ values are uniformly characterized by late 10% rise times, indicating that in these model realizations downwelling Xe concentrations remain very low throughout most of Earth history. **c**, Variation in sigmoidal growth rates (parameter α) allows testing of near-linear (low α , sampling for a limited time interval) or near-step (high α) functions. Near-linear model realizations have a time of 50% rise that is about five times the time of 10% rise (dashed light-grey line), whereas step functions approach a 1:1 line (solid dark-grey line). Successful regassing histories with late times of 10% rise are characterized by rapid growth, approaching a step function, to a relatively high present-day $^{130}\text{Xe}_d$.



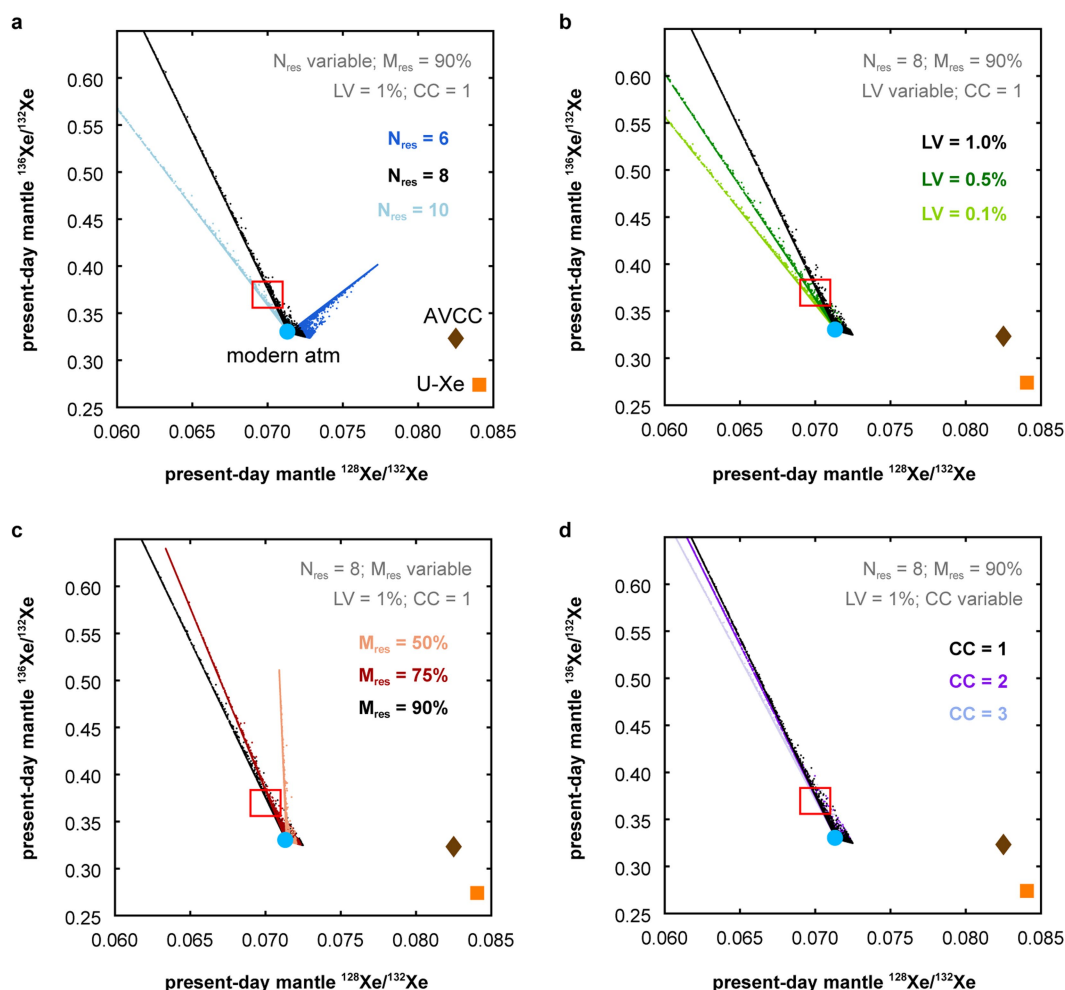
Extended Data Fig. 5 | Successful regassing histories for varying model parameters. **a–d**, To test model sensitivity to the input parameters, we vary the number of mantle reservoir masses processed (N_{res}), convecting mantle reservoir mass (M_{res}), initial ^{130}Xe concentration (LV) and continental crust model (CC), and collect all successful $^{130}\text{Xe}_d$ (**a**, **c**) and mantle ^{130}Xe concentrations (**b**, **d**) over time. For sigmoidal $^{130}\text{Xe}_d$, solutions are found for $N_{\text{res}} = \{5, 6, 7, 8, 9\}$, M_{res} of 50%, 75% and 90% of the whole mantle mass, LV of 0.1%, 0.5% and 1% chondritic late veneers, and all three CCs. Extended Data Figs. 6, 7 illustrate trade-offs between

individual parameters; for instance, high N_{res} values generate solutions only with high LV. For all sigmoidal solutions, regassing is limited early in Earth history, and the mantle shifts from net degassing to net regassing after ~ 2.5 Gyr ago. For exponential $^{130}\text{Xe}_d$, solutions are found for $N_{\text{res}} = \{5, 6, 7, 8, 9\}$, M_{res} of 50%, 75% and 90% of the whole mantle mass, LV of 0.1%, 0.5% and 1% chondritic late veneers, and all three CCs. For all solutions, regassing is limited early in Earth history, and the mantle shifts from net degassing to net regassing after ~ 2.5 Gyr ago.



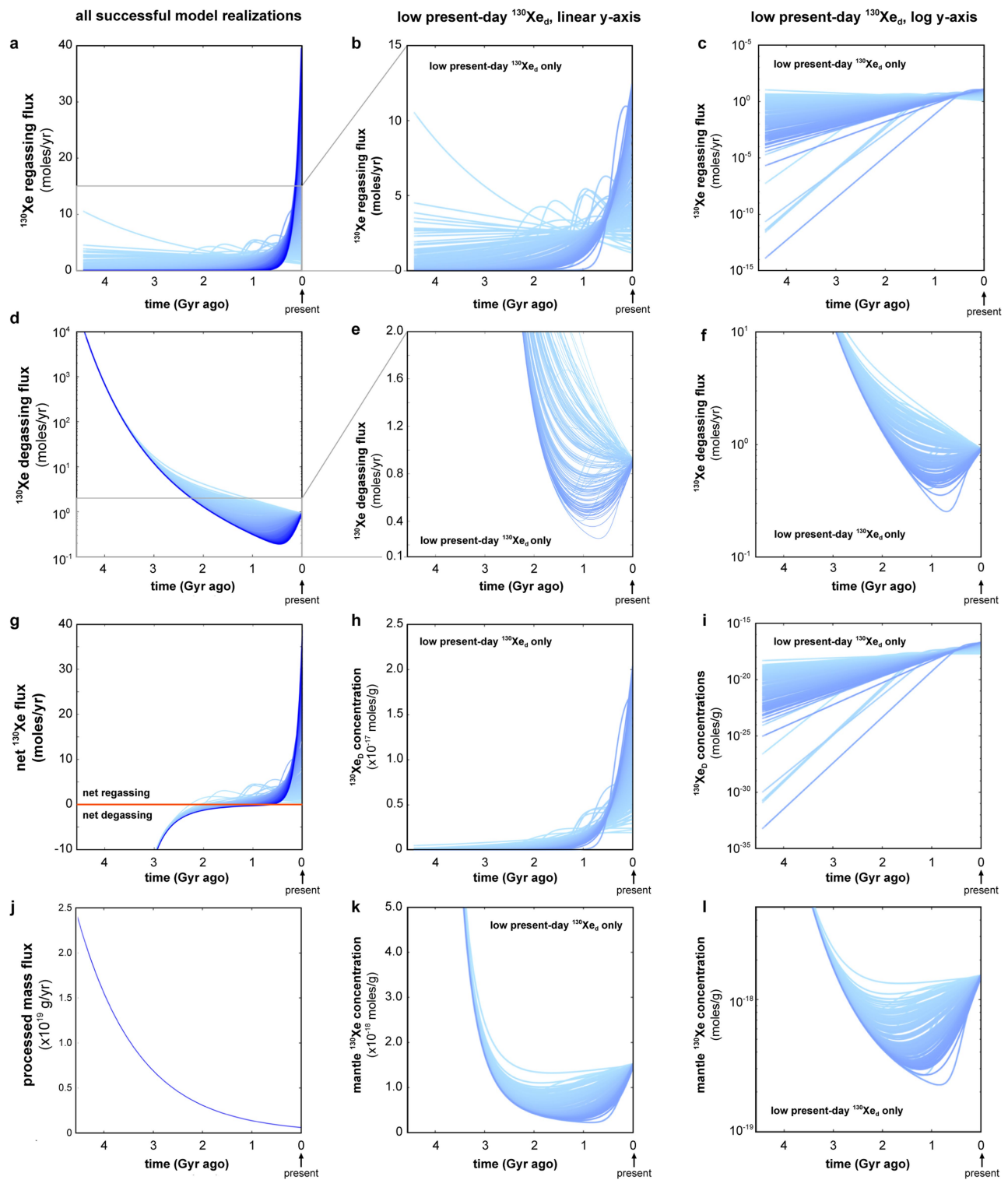
Extended Data Fig. 6 | Sensitivity of ^{130}Xe and $^{128}\text{Xe}/^{130}\text{Xe}$ to model parameters. Present-day mantle ^{130}Xe concentration and the ratio of two primordial stable isotopes, ^{128}Xe and ^{130}Xe are shown for different model parameter combinations. Four parameters are explored: those affecting the mantle processing-rate history (M_{res} and N_{res}), LV (initial ^{130}Xe concentrations corresponding to a late veneer fraction between 0.1% and 1%) and CC (Extended Data Fig. 3). In each panel, three of these parameters are held constant and the other is varied to illustrate model sensitivity to the varied parameter. Each cloud of points represents the range of present-day ^{130}Xe and $^{128}\text{Xe}/^{130}\text{Xe}$ generated by different regassing histories for the specified N_{res} , M_{res} , LV and CC. The red rectangle indicates the estimated present-day mantle ^{130}Xe concentration and $^{128}\text{Xe}/^{130}\text{Xe}$ range. Dots that fall within the red rectangle represent the family of regassing histories that successfully reproduce the present-day mantle composition for each parameter combination. The reference case shown in Figs. 3, 4 ($M_{\text{res}} = 90\%$, $N_{\text{res}} = 8$, LV = 1%, CC = 1) is shown as

a cloud of black points in all panels. **a**, A higher mantle processing rate ($N_{\text{res}} = 10$) results in low ^{130}Xe concentrations for successful $^{128}\text{Xe}/^{130}\text{Xe}$ ratios, and $^{128}\text{Xe}/^{130}\text{Xe}$ ratios that are too low for successful ^{130}Xe concentrations. **b**, Higher late-veneer fractions correspond to higher initial ^{130}Xe concentrations in the mantle. For the same mantle processing-rate history, LV = 0.1% yields present-day mantle ^{130}Xe concentrations that are too low given successful $^{128}\text{Xe}/^{130}\text{Xe}$ ratios. The effect of low LV can be offset by lowering N_{res} and thus decreasing the total amount of degassing over Earth history; thus, N_{res} and LV can be co-varied to find solutions. **c**, The effect of M_{res} is minimal because degassing is parameterized through the number of reservoir masses processed over Earth history. Some difference is evident at high present-day mantle ^{130}Xe abundances because the same $^{130}\text{Xe}_d$ regassing rate parameter space is explored against different absolute degassing rates. **d**, The continental crust model has no effect on budgets of primordial Xe isotopes.



Extended Data Fig. 7 | Sensitivity of fissiogenic Xe to model parameters. Present-day outcomes are shown in ^{128}Xe – ^{132}Xe – ^{136}Xe isotopic space for different model parameter combinations. Four parameters are explored: parameters affecting the mantle processing-rate history (M_{res} and N_{res}), the initial mantle ^{130}Xe concentration ($\text{LV} = 0.1\%–1\%$), and CC (Extended Data Fig. 3). In each panel, three of these parameters are held constant and the other is varied to illustrate model sensitivity to the varied parameter. Each cloud of points represents the range of present-day $^{128}\text{Xe}/^{132}\text{Xe}$ and $^{136}\text{Xe}/^{132}\text{Xe}$ generated by different regassing histories given the specified N_{res} , M_{res} , LV and CC . The red rectangle indicates the estimated present-day mantle $^{128}\text{Xe}/^{132}\text{Xe}$ and $^{136}\text{Xe}/^{132}\text{Xe}$ range. Dots that fall within the red rectangle represent the family of regassing histories that successfully reproduce present-day

mantle composition for each parameter combination. The reference case shown in the main-text figures ($M_{\text{res}} = 90\%$, $N_{\text{res}} = 8$, $\text{LV} = 1\%$, $\text{CC} = 1$) is shown as a cloud of black points in all panels. The orange square is U-Xe, the brown diamond is average carbonaceous chondrites (AVCC) and the blue circle is the modern atmosphere. **a**, Higher mantle processing rates push present-day compositions towards fissiogenic Xe components. **b**, Lower late-veneer fractions correspond to present-day compositions closer to fissiogenic Xe components. **c**, A relatively low mass of the convecting mantle means that the mantle must be more depleted in U to satisfy mass balance with the continental crust (Methods). Thus, for low M_{res} , the impact of fission is muted compared to high M_{res} . **d**, The continental crust model has a limited effect on present-day Xe isotopic compositions.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | ^{130}Xe fluxes over time in successful model realizations. **a–l**, Regassing fluxes (**a–c**), degassing fluxes (**d–f**), net fluxes (**g**), $^{130}\text{Xe}_d$ concentrations (**h, i**), mass flux (**j**) and mantle ^{130}Xe concentrations (**k, l**) are illustrated for an initial mantle ^{130}Xe concentration of 3.2×10^8 atoms per gram ($LV = 1\%$), a convecting mantle reservoir that is 90% of the mass of the whole mantle, and 8 mantle reservoir masses processed over Earth history. Fluxes are reported in moles per year and concentrations are reported in moles per gram. Panels in the left column show results from all successful model realizations (same results as those shown in Figs. 3, 4) and illustrate the ^{130}Xe regassing flux (**a**), ^{130}Xe degassing flux (**d**), ^{130}Xe net flux (**g**) and mass flux over time (**j**). Panels in the central column show zoomed-in windows with only low- $^{130}\text{Xe}_d$ successful model realizations (light-blue lines), as these largely overlap with each other and are difficult to resolve in the full-scale panels. The right column replicates the central column with semi-logarithmic axes. The regassing ^{130}Xe flux time series (**a–c**) is

the product of the downwelling mass flux time series (**j**; exponentially decreasing with time) and the $^{130}\text{Xe}_d$ concentration over time (sigmoidally increasing; **h, i**). Time series for ^{130}Xe regassing fluxes with high present-day $^{130}\text{Xe}_d$ (darkest-blue lines in **a**) start near zero owing to near-zero ^{130}Xe concentrations and then rapidly rise as the $^{130}\text{Xe}_d$ concentration increases faster than the modest decline in mass flux later in Earth history. ^{130}Xe flux time series with low present-day $^{130}\text{Xe}_d$ (lightest-blue lines in **a–c**) start a protracted, low-magnitude rise relatively early in Earth history. These translate to regassing flux time series that start near zero, rise and then decline with the exponentially decreasing mass flux (**b, c**). Time series for ^{130}Xe degassing fluxes (**d–f**) are the product of the downwelling mass flux time series (**j**) and the mantle ^{130}Xe concentration over time (**k, l**), which responds to both degassing and regassing. The net flux over time (**g**) is the difference between the regassing flux and degassing flux at any given time. The mantle shifts from net degassing to net regassing at some time after 2.5 Gyr ago.

Extended Data Table 1 | Notation

Subscripts and superscripts	
d	in downwellings
m	in convecting mantle (mid-ocean ridge basalt source)
p	primordial
r	regassed
Regassing history	
$^{130}\text{Xe}_d$	concentration of ^{130}Xe in downwellings over time (atoms/g)
<i>Sigmoidal $^{130}\text{Xe}_d$ parameters</i>	
$^{130}\text{Xe}_d^k$	$^{130}\text{Xe}_d$ sigmoid carrying capacity (0 to 5×10^8 atoms/g)
α	sigmoid growth rate (10^{-10} to 10^{-8} Gyr $^{-1}$)
β	sigmoid inflection time (0.08 - 10 Gyr)
<i>Exponential $^{130}\text{Xe}_d$ parameters</i>	
$^{130}\text{Xe}_d^{\text{final}}$	$^{130}\text{Xe}_d$ in the present day (0 to 5×10^8 atoms/g)
τ	$^{130}\text{Xe}_d$ time constant (10^{-11} to 5×10^{-8} Gyr $^{-1}$)
Mantle processing and fissiogenic ingrowth	
LV	late veneer fraction (0.1%, 0.5%, 1% Earth mass)
M_{res}	mass of convecting mantle reservoir (50%-90%, 2×10^{27} to 3.6×10^{27} g)
N_{res}	number of reservoir masses processed over Earth history
Q	mantle processing rate (g/yr)
Q_p	present-day mantle processing rate (6.1×10^{17} g/yr)
η	processing rate time constant
T	age of the Earth (4.568 Gyr)
t	time (yr)
dM	mass of mantle processed in a time step (g)
dU _{CC}	change in ^{238}U concentration in convecting mantle in a time step (atoms/g)
CC	continental crust growth model (1,2,3)
λ_{244}	^{244}Pu decay constant (8.6643×10^{-9} yr $^{-1}$)
λ_{238}	^{238}U decay constant (1.5514×10^{-10} yr $^{-1}$)
Y_{Pu}^{136}	Fission yield of ^{136}Xe from ^{244}Pu (7×10^{-5})
Y_{U}^{136}	Fission yield of ^{136}Xe from ^{238}U (3.43×10^{-8})
ψ	Fission Xe isotope mass (131, 132, 134, 136)

Extended Data Table 2 | Xe isotopic compositions

ratio	Present-day convecting mantle*		Fission Xe	
	mantle min	mantle max	²⁴⁴ Pu†	²³⁸ U‡
¹²⁸ Xe/ ¹³⁰ Xe	0.475	0.478	n/a	n/a
¹²⁸ Xe/ ¹³² Xe	0.069	0.071	0	0
¹³⁰ Xe/ ¹³² Xe	0.1445	0.1493	0	0
¹³¹ Xe/ ¹³² Xe	0.7608	0.7786	0.1449	0.2777
¹³⁴ Xe/ ¹³² Xe	0.4082	0.4302	1.437	1.041
¹³⁶ Xe/ ¹³² Xe	0.3559	0.3835	1.738	1.120

min, minimum; max, maximum; n/a, not applicable.

*Limits derived from refs ^{11,14,16,23}.

†Error-weighted average of data from refs ^{46,56,57}.

‡Error-weighted average of data from refs ^{58–61}.

Shared evolutionary origin of vertebrate neural crest and cranial placodes

Ryoko Horie^{1,4}, Alex Hazbun^{1,2,4}, Kai Chen^{1,4}, Chen Cao¹, Michael Levine^{1,2*} & Takeo Horie^{1,3*}

Placodes and neural crests represent defining features of vertebrates, yet their relationship remains unclear despite extensive investigation^{1–3}. Here we use a combination of lineage tracing, gene disruption and single-cell RNA-sequencing assays to explore the properties of the lateral plate ectoderm of the proto-vertebrate, *Ciona intestinalis*. There are notable parallels between the patterning of the lateral plate in *Ciona* and the compartmentalization of the neural plate ectoderm in vertebrates⁴. Both systems exhibit sequential patterns of *Six1/2*, *Pax3/7* and *Msx* expression that depend on a network of interlocking regulatory interactions⁴. In *Ciona*, this compartmentalization network produces distinct but related types of sensory cells that share similarities with derivatives of both cranial placodes and the neural crest in vertebrates⁵. Notably, bipolar tail neurons were readily transformed into palp sensory cells, a proto-placodal sensory cell type that arises from the anterior-most regions of the lateral plate in the *Ciona* tadpole⁶. Proof of transformation was confirmed by whole-embryo single-cell RNA-sequencing assays. These findings suggest that compartmentalization of the lateral plate ectoderm preceded the advent of vertebrates, and served as a common source for the evolution of both cranial placodes and neural crest^{3,4}.

Placodes and neural crest are the key ontogenetic novelties underlying vertebrate cephalization^{1–3}. However, their evolutionary origins remain uncertain despite recent evidence that invertebrate chordates contain rudiments of both cell types^{4–14}. Here we obtain a more comprehensive view of the lateral plate ectoderm in *Ciona*, because it is the source of the cell types that are related to placodal and neural crest derivatives in vertebrates.

Lineage-tracing methods were used to identify four derivatives of the lateral plate ectoderm: palp sensory cells (PSCs) arising from the a8.20 and a8.18 lineages¹², anterior apical trunk epidermal neurons (aATENs) (a8.26 lineage)¹⁴, posterior apical trunk epidermal neurons (pATENs) (b8.20 lineage)⁹ and bipolar tail neurons (BTNs; b8.18 lineage)⁵ (Fig. 1a, b and Extended Data Fig. 1). The aATENs were previously shown to possess dual properties of placode-derived chemosensory neurons (for example, olfactory neurons) and GnRH-expressing neurosecretory neurons (for example, hypothalamic GnRH neurons)¹⁴. BTNs are thought to share properties with neural-crest-derived dorsal root ganglia⁵. Analysis of the regulatory ‘blueprint’ of the *Ciona* embryo identified several determinants of the lateral plate ectoderm, including *Dmrt.a*, *Foxc*, *Six1/2*, *Pax3/7* and *Msx*¹⁵. *Dmrt.a*, *Foxc* and *Six1/2* are expressed in the anterior-most regions (a8.20 and a8.18 and a8.26 lineages)^{12,14,16} (Fig. 1c, d), whereas *Msx* is selectively expressed in posterior regions (b8.20 and b8.18 lineages)^{15,17} (Fig. 1d). *Pax3/7* is found in both anterior and posterior regions, spanning a8.26, b8.20 and b8.18 lineages¹⁸ (Fig. 1c).

We obtained evidence for interlocking regulatory interactions among these lateral plate determinants (Fig. 2a–d and Extended

Data Figs. 2–5). *Dmrt.a* activates *Foxc* and *Six1/2* expression in the anterior-most regions^{15,16} (Fig. 2a, b). There is expansion of *Six1/2* and *Eya* expression in the tail regions of *Msx* morpholino antisense

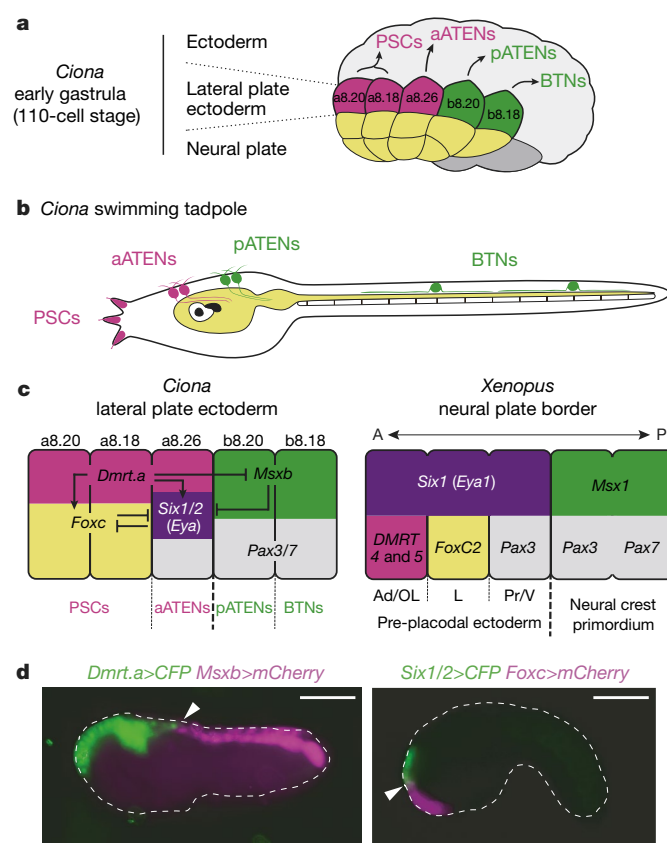


Fig. 1 | Lateral plate ectoderm. **a**, Summary of lateral plate derivatives at early gastrula (110-cell stage). Magenta, progenitors of PSCs (a8.18 and a8.20 lineage) and aATENs (a8.26 lineage) lineages. Green, pATENs and BTNs arising from b8.20 and b8.18 lineages, respectively. Yellow, neural plate. **b**, Diagram of *Ciona* tadpole showing the position of PSCs, aATENs, pATENs and BTNs. **c**, Summary of *Ciona* lateral plate ectoderm and *Xenopus* neural plate border. Magenta, *Dmrt.a*-expressing blastomeres (a8.20, a8.18 and a8.26 lineage); green, *Msx*-expressing blastomeres (b8.20 and b8.18 lineage); purple, prospective *Six1/2*- and *Eya*-expressing blastomeres (a8.26 lineage); yellow, *Foxc*-expressing blastomeres (a8.20 and a8.18 lineage); grey, *Pax3/7*-expressing blastomeres (a8.26, b8.20 and b8.18 lineage). A, anterior; Ad, adenohypophyseal placode; L, lens placode; OL, olfactory placode; P, posterior; Pr, profundal placode; V, trigeminal placode. **d**, Left, tailbud embryo injected with *Dmrt.a* > CFP (green) and *Msx* > mCherry (magenta) reporter genes. The arrowhead indicates the boundary that separates the regions in which *Dmrt.a* and *Msx* is expressed. Right, tailbud embryo injected with *Six1/2* > CFP and *Foxc* > mCherry reporter genes. The arrowhead indicates the boundary that separates the regions in which *Six1/2* and *Foxc* is expressed. Anterior is to the left. Scale bars, 100 μm.

¹Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. ²Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ³Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka, Japan. ⁴These authors contributed equally: Ryoko Horie, Alex Hazbun, Kai Chen. *e-mail: msl2@princeton.edu; horie@shimoda.tsukuba.ac.jp

oligonucleotide (MO) morphants, suggesting that *Mxsb* functions as a repressor to delineate the trunk–tail boundary of the lateral plate ectoderm (Fig. 2a–c and Extended Data Fig. 2). The characterization of a minimal *Six1/2* enhancer is consistent with direct repression by *Mxsb* (Extended Data Fig. 4). Furthermore, there is an anterior expansion of *Mxsb* expression in *Dmrt.a* MO morphant mutants, raising the possibility of reciprocal repression of *Mxsb* by either *Six1/2* (and *Eya*) or *Dmrt.a* (Extended Data Fig. 5).

There are notable parallels in the compartmentalization of the lateral plate in vertebrates and ascidians (Fig. 1c). In *Xenopus*, *Dmrt.a* homologues 4 and 5 specify the adenohypophyseal and olfactory placodes within anterior regions of the pan-placodal primordium^{19,20}. They do not appear to regulate *Six1* as was seen in ascidians. Nonetheless, in *Ciona* *Dmrt.a* gives way to sequential expression of *Six1/2* and *Mxsb* at the trunk–tail boundary, similar to that seen in vertebrates (Fig. 1c, d). Moreover, the overlapping patterns of *Six1/2* and *Pax3/7* expression seen in the *Ciona* aATEN lineage are similar to the patterns that delineate specific compartments within the pan-placodal primordium in vertebrates^{21–23}.

The most pronounced deviation between the *Ciona* and vertebrate regulatory fate maps is the compartmentalization of the *Ciona* anterior lateral plate into two distinct domains that showed mutually exclusive expression of *Foxc* (PSCs) and *Six1/2* (aATENs) (Fig. 1c, d). We determined whether these territories might share common developmental properties, because vertebrate orthologues of *Foxc* have been implicated in delineating placodal derivatives such as the eye lens^{24–26}. *Foxc* morphants were obtained by injection of a sequence-specific MO that targets the translation start site of the endogenous *Foxc* gene. They show a loss of gene expression of PSC markers (Fig. 2e, f) as well as an unexpected phenotype: ectopic expression of a *Six1/2* reporter gene in palp regions that produced PSCs (Fig. 2d, g, h and Extended Data Fig. 6). Thus, *Foxc* appears to function as a key determinant of PSC identity by activating PSC markers and inhibiting an alternative aATEN identity. This transformation of PSCs into aATENs suggests that the *Foxc* and *Six1/2* territories of the anterior lateral plate use a similar developmental program for specifying sensory cells.

It has been suggested that BTNs are related to dorsal root ganglia, which are derived from neural crest cells in vertebrates⁵. This observation raises the possibility that tail regions of the lateral plate ectoderm possess ‘proto-neural crest’ properties⁵, possibly indicating a common origin of cranial placodes and the neural crest from lateral plate ectoderm¹. To explore this possibility, we investigated whether BTNs could be transformed into other derivatives of the lateral plate ectoderm. We misexpressed *Foxc* in posterior regions of the lateral plate using *Pax3/7* and *Mxsb* regulatory sequences. Mutant tailbud embryos showed variable transformations of BTNs into PSCs (Fig. 2i–k) without changes in other tail structures (for example, neural tube and notochord). Some BTNs expressed only PSC marker genes (for example, $\beta\gamma$ -crystallin; arrow, Fig. 2j) whereas others expressed both PSC and BTN markers (for example, *Asic1b*; arrowhead, Fig. 2j, k). These observations indicate that anterior and posterior regions of the lateral plate ectoderm have a similar developmental program for the specification of related but distinct sensory cell types. It is therefore possible that the entire lateral plate of the last shared tunicate and vertebrate ancestor is the source of both placodal and neural crest derivatives in vertebrates.

Because the transformation of BTNs into supernumerary PSCs is pivotal to our proposal that the compartmentalized lateral plate ectoderm produces distinct but related sensory cell types, we used single-cell RNA-sequencing (RNA-seq) assays to characterize transformed BTNs by taking advantage of the well-defined lineages and the fact that *Ciona* tailbud embryos consist of a small number of cells (around 1,500 cells). Embryos were injected with the *Pax3/7 > Foxc* transgene (Fig. 2j), grown to the late tailbud stage, dissociated and sequenced using the 10 \times microfluidics platform. Approximately 5,000 cells were sequenced in order to obtain sufficient coverage (around 3 \times) to ensure reliable detection of PSCs, BTNs and transformed cell types. Unequivocal identification of cells expressing the *Pax3/7* transgene was

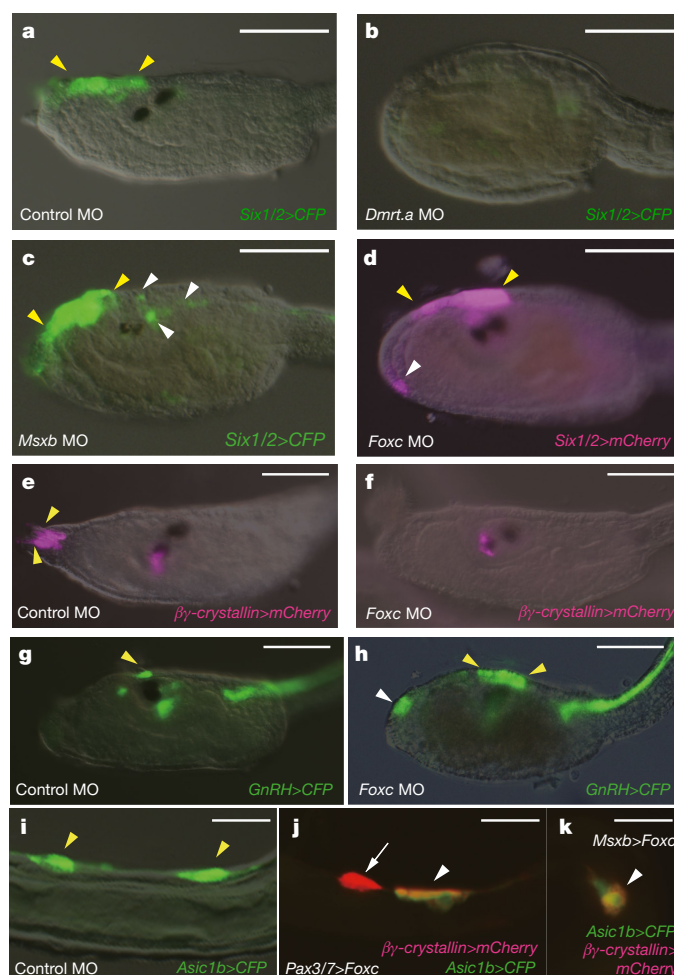


Fig. 2 | Functional analysis of the lateral plate ectoderm. a–c, Head regions of larvae that were injected with a *Six1/2 > CFP* reporter gene. a, *Six1/2* expression in the proto-placodal region of control MO injected tadpoles (49 of 49 larvae displayed this expression pattern). The yellow arrowheads identify the normal location of *Six1/2* expression. b, Loss of expression in *Dmrt.a* morphants (49 of 49 larvae). c, Expanded expression of the *Six1/2 > CFP* reporter gene (white arrowheads) in *Mxsb* morphants (42 of 50 larvae showed this expansion pattern). d, Head regions of a larva that was injected with *Foxc* MO and *Six1/2 > mCherry* reporter gene. There is ectopic expression (white arrowhead) in the palp regions of *Foxc* morphants (35 of 47 larvae showed this phenotype). e, f, Larvae injected with $\beta\gamma$ -crystallin $>$ mCherry reporter gene. Yellow arrowheads indicate the $\beta\gamma$ -crystallin expressing PSCs in control MO injected larvae (51 of 51 larvae display this expression pattern). f, There is a loss of these cells in *Foxc* morphants (108 of 108 larvae showed this phenotype). g, h, Larvae injected with a *GnRH > CFP* reporter gene. Yellow arrowheads identify the *GnRH* expressing aATENs in a control larva (59 of 59 larvae displayed expression in aATENs). h, There is ectopic expression in the palp regions of *Foxc* morphants (white arrowheads) (28 of 40 injected larvae showed this phenotype). i–k, Tail regions of larvae injected with *Asic1b > CFP* (i) and also injected with $\beta\gamma$ -crystallin $>$ mCherry reporter gene (j, k). Yellow arrowheads identify the *Asic1b* expressing BTNs in a control larva (83 of 83 larvae displayed this phenotype). j, Ectopic expression of the $\beta\gamma$ -crystallin $>$ mCherry reporter gene in tail regions (white arrowheads) upon misexpression of *Foxc* by the *Pax3/7* enhancer (26 of 55 larvae showed this phenotype). k, Same as j except that *Mxsb* regulatory sequences were used to misexpress *Foxc* (31 of 57 larvae showed misexpression of $\beta\gamma$ -crystallin $>$ mCherry). Anterior to the left; scale bars, 100 μ m (a–h), 20 μ m (i–k).

provided by the insertion of a unique 450-bp sequence tag positioned downstream of the *Foxc* coding region.

t-distributed stochastic neighbour embedding (*t*-SNE) projections reveal 20 cell clusters that represent different tissues, including

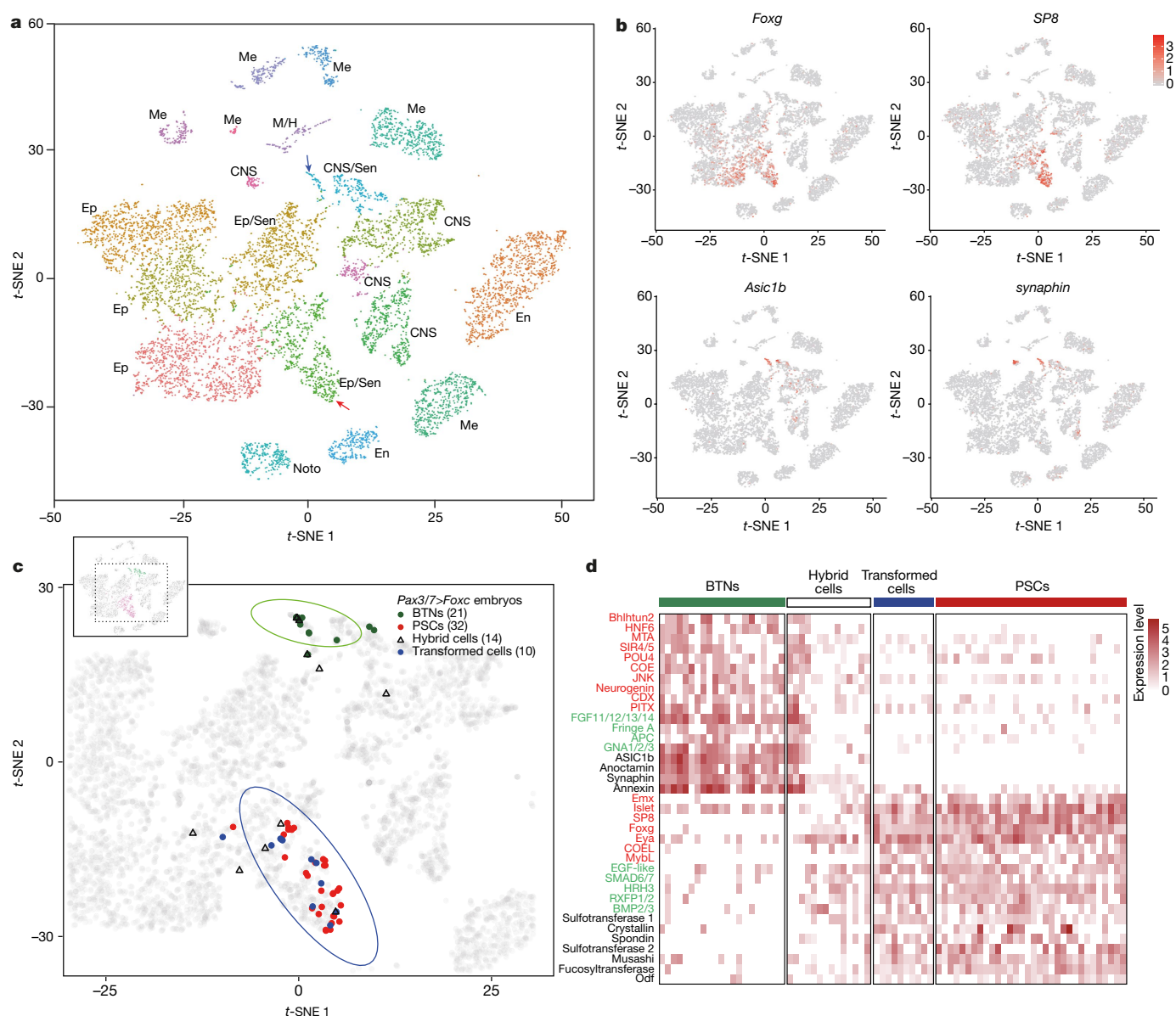


Fig. 3 | Single-cell RNA-seq analysis of BTN transformations. **a**, The *t*-SNE projection map of dissociated cells from wild-type and mutant tailbud-stage embryos that misexpress *Foxc* in tail regions using *Pax3/7* regulatory DNAs (*Pax3/7* > *Foxc* transgene). Each dot corresponds to the transcriptome of a single cell, and cells possessing similar transcriptome profiles map near each other. All of the major tissue types in tailbud-stage embryos were identified. CNS, central nervous system; Ep, epidermis; En, endoderm; M/H, heart and muscle; Me, mesenchyme; Noto, notochord; sen, sensory neurons. Identification is based on the expression of known marker genes (Extended Data Fig. 8b, and Supplementary Table 1). Red arrow identifies PSCs. **b**, Distribution of marker genes expressed in PSCs

notochord, endoderm, tail muscles, mesenchyme, epidermis and central nervous system (Fig. 3a, Extended Data Fig. 7a, b and Supplementary Table 1). BTNs were identified by their expression of key marker genes, such as *Asic1b* and *synaphin*, whereas PSCs expressed a distinct set of markers, including *islet*, *SP8* and *Foxg* (Fig. 3b, Extended Data Fig. 8 and Supplementary Table 1). Transformed BTNs were defined as those expressing the *Pax3/7* transgene, lacking expression of BTN markers (*Asic1b* and *synaphin*), acquiring expression of PSC markers (*islet*, *Foxg* and *SP8*) and clustering within the 95% confidence interval ellipse of native PSCs (bottom oval, Fig. 3c). Partially transformed BTNs were defined as those expressing *Pax3/7* transgenes, lacking expression of only one of the BTN marker genes and acquiring expression of only a

(*Foxg* and *SP8*) and BTNs (*Asic1b* and *synaphin*) within *t*-SNE projections as shown in (a). **c**, Distribution of cells expressing transgenes, which identifies cells that misexpress the PSC determinant, *Foxc*. BTNs (dark green dots; *n* = 21), PSCs (red dots; *n* = 32), hybrid cells that express both PSC and BTN marker genes (light blue triangles; *n* = 14), and transformed cells that express PSC markers (blue dots; *n* = 10). The grey dots (*n* = 10,103) correspond to all dissociated cells that were sequenced in these experiments. **d**, Heat map of BTNs, PSCs, transformed cells and hybrid cells showing the relative expression of a select group of genes encoding transcription factors (red), signalling components (green) and cellular effectors (black).

subset of PSC markers. The transcriptomes of hybrid cells tended to map outside of the 95% confidence interval ellipse of PSCs (Fig. 3c). Altogether, 45 BTNs were identified in the whole-embryo single-cell transcriptome datasets. About half (21) were untransformed and displayed the native BTN transcriptome profile, whereas the other half were either fully transformed into PSCs (10) or partially transformed (14) into a hybrid BTN–PSC identity (Fig. 3d). These findings closely mirror the direct visualization of reporter gene expression in transgenic embryos, in which BTNs exhibit variable expression of $\beta\gamma$ -crystallin and *Asic1b* reporter genes (Fig. 2j, k).

The *Pax3/7* > *Foxc* transgene is expressed in the lateral plate and additional tissues, such as the mesenchyme, which is a common

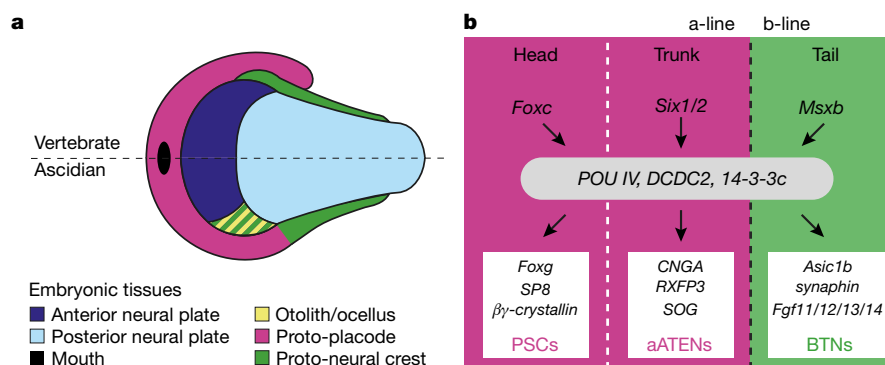


Fig. 4 | Compartmentalization of *Ciona* lateral plate ectoderm.

a, Schematic illustration of the neural plate and its lateral border in vertebrate (top half) and ascidians (bottom half). We propose that both proto-placode and proto-neural crest evolved from the entire lateral plate in the last shared tunicate-vertebrate ancestor. Otolith/ocellus contribution to the proto-neural crest is based on a previous publication²⁹. **b**, Provisional gene regulatory network for the compartmentalization,

specification and differentiation of the lateral plate ectoderm into sensory cell types in *C. intestinalis*. During early embryogenesis, *Foxc* and *Six1/2* determine the head-trunk boundary (white dotted line) of proto-placode lineages (magenta) and *Mxsb* patterns the proto-neural crest territory (green). In tailbud stages, a common developmental program specifies sensory progenitors within each compartment. *CNGA*, *RXFP3* and *SOG* (also known as *chemokine-like*) expression is based on a previous study¹⁴.

site of ectopic expression of *Ciona* transgenes²⁷. These other sites of *Foxc* expression do not undergo transformation in cell identity, but instead show native transcriptome profiles (Extended Data Fig. 7e, f). Altogether, the single-cell RNA-seq assays strengthen the evidence that BTNs are transformed into PSCs, suggesting the use of a similar developmental program for the specification of different sensory cells arising from head, trunk and tail regions of the lateral plate ectoderm.

We present evidence that the antero-posterior compartmentalization of the *Ciona* lateral plate leads to the development of related but distinct sensory cell types (Fig. 4). PSCs, aATENs and BTNs express a common suite of regulatory genes and cell identity genes (for example, *POU IV*, *DCDC2* and *14-3-3c*), despite their different origins along the lateral plate (Extended Data Fig. 9). *Foxc*, *Six1/2* and *Mxsb* impose distinctive signatures of gene activity, leading to the specification of diverse sensory cell types. There are notable parallels with the regional specification of distinct somatosensory neurons arising from placodal and neural crest territories in vertebrates²⁸. We therefore suggest that a compartmentalized lateral plate preceded the advent of vertebrates, and served as a common source for the evolution of both cranial placodes and neural crest.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0385-7>.

Received: 18 August 2017; Accepted: 13 June 2018;

Published online 1 August 2018.

- Northcutt, R. G. & Gans, C. The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. *Q. Rev. Biol.* **58**, 1–28 (1983).
- Baker, C. V. & Bronner-Fraser, M. The origins of the neural crest. Part II: an evolutionary perspective. *Mech. Dev.* **69**, 13–29 (1997).
- Schlosser, G. Do vertebrate neural crest and cranial placodes have a common evolutionary origin? *BioEssays* **30**, 659–672 (2008).
- Schlosser, G., Patthey, C. & Shimeld, S. M. The evolutionary history of vertebrate cranial placodes II. Evolution of ectodermal patterning. *Dev. Biol.* **389**, 98–119 (2014).
- Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371–374 (2015).
- Wagner, E., Stolfi, A., Gi Choi, Y. & Levine, M. Islet is a key determinant of ascidian palp morphogenesis. *Development* **141**, 3084–3092 (2014).
- Manni, L. et al. Neurogenic and non-neurogenic placodes in ascidians. *J. Exp. Zool. B Mol. Dev. Evol.* **302B**, 483–504 (2004).
- Mazet, F. et al. Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. *Dev. Biol.* **282**, 494–508 (2005).
- Pasini, A. et al. Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system. *PLoS Biol.* **4**, e225 (2006).

- Imai, J. H. & Meinertzhagen, I. A. Neurons of the ascidian larval nervous system in *Ciona intestinalis*: II. Peripheral nervous system. *J. Comp. Neurol.* **501**, 335–352 (2007).
- Horie, T., Kusakabe, T. & Tsuda, M. Glutamatergic networks in the *Ciona intestinalis* larva. *J. Comp. Neurol.* **508**, 249–263 (2008).
- Wagner, E. & Levine, M. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012).
- Patthey, C., Schlosser, G. & Shimeld, S. M. The evolutionary history of vertebrate cranial placodes—I: cell type evolution. *Dev. Biol.* **389**, 82–97 (2014).
- Abitua, P. B. et al. The pre-vertebrate origins of neurogenic placodes. *Nature* **524**, 462–465 (2015).
- Imai, K. S., Levine, M., Satoh, N. & Satou, Y. Regulatory blueprint for a chordate embryo. *Science* **312**, 1183–1187 (2006).
- Tresser, J. et al. *doublesex/mab3 related-1 (dmrt1)* is essential for development of anterior neural plate derivatives in *Ciona*. *Development* **137**, 2197–2203 (2010).
- Aniello, F. et al. Identification and developmental expression of *Ci-msxb*: a novel homologue of *Drosophila msh* gene in *Ciona intestinalis*. *Mech. Dev.* **88**, 123–126 (1999).
- Wada, H., Holland, P. W., Sato, S., Yamamoto, H. & Satoh, N. Neural tube is partially dorsalized by overexpression of *HrPax-37*: the ascidian homologue of *Pax-3* and *Pax-7*. *Dev. Biol.* **187**, 240–252 (1997).
- Huang, X., Hong, C. S., O'Donnell, M. & Saint-Jeannet, J. P. The doublesex-related gene, *Xdmt4*, is required for neurogenesis in the olfactory system. *Proc. Natl Acad. Sci. USA* **102**, 11349–11354 (2005).
- Parlier, D. et al. The *Xenopus* doublesex-related gene *Dmrt5* is required for olfactory placode neurogenesis. *Dev. Biol.* **373**, 39–52 (2013).
- Ahrens, K. & Schlosser, G. Tissues and signals involved in the induction of placodal *Six1* expression in *Xenopus laevis*. *Dev. Biol.* **288**, 40–59 (2005).
- Schlosser, G. & Ahrens, K. Molecular anatomy of placode development in *Xenopus laevis*. *Dev. Biol.* **271**, 439–466 (2004).
- Pieper, M., Eagleson, G. W., Wosniok, W. & Schlosser, G. Origin and segregation of cranial placodes in *Xenopus laevis*. *Dev. Biol.* **360**, 257–275 (2011).
- Köster, M., Dillinger, K. & Knöchel, W. Activin A signaling directly activates *Xenopus* winged helix factors *XFD-4/4'*, the orthologues to mammalian *MFH-1*. *Dev. Genes Evol.* **210**, 320–324 (2000).
- Bailey, A. P., Bhattacharyya, S., Bronner-Fraser, M. & Streit, A. Lens specification is the ground state of all sensory placodes, from which FGF promotes olfactory identity. *Dev. Cell* **11**, 505–517 (2006).
- Berry, F. B. et al. Functional interactions between *FOXC1* and *PITX2* underlie the sensitivity to *FOXC1* gene dose in Axenfeld-Rieger syndrome and anterior segment dysgenesis. *Hum. Mol. Genet.* **15**, 905–919 (2006).
- Corbo, J. C., Levine, M. & Zeller, R. W. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589–602 (1997).
- Schlosser, G. Induction and specification of cranial placodes. *Dev. Biol.* **294**, 303–351 (2006).
- Abitua, P. B., Wagner, E., Navarrete, I. A. & Levine, M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature* **492**, 104–107 (2012).

Acknowledgements We thank all members of the LSI genome facility for technical support of the single-cell RNA-seq assays and analysis; R. Yoshida and C. Imaizumi and all other members of staff at the Maizuru Fisheries Research Station of Kyoto University for providing *Ciona intestinalis*. This study was supported by a grant from the NIH to M.L. (NS076542) and by Grants-in-Aid for Scientific Research from JSPS to T.H. (24687008, 16K07433). T.H. was supported by the Pre-Strategic Initiatives, University of Tsukuba. A portion of *Ciona intestinalis* and plasmids used in this study were provided by the

National Bio-Resource Project (NBRP) of the MEXT, Japan. A.H. was partially supported by a fellowship from the Colombian Government (Colciencias 568).

Reviewer information *Nature* thanks N. Satoh, G. Schlosser, S. Shimeld and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.H., A.H. and M.L. conceived the project. T.H., K.C. and M.L. designed the experiments. R.H., K.C. and T.H. performed the experiments. R.H., A.H., K.C., C.C., T.H. and M.L. analysed and interpreted the data. T.H., A.H., K.C., C.C. and M.L. wrote the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0385-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0385-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to M.L. or T.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Biological materials. Wild-type *C. intestinalis* adults were obtained from M-Rep and the National Bio-Resource Project for *Ciona* in Japan. Sperm and eggs were collected by dissecting the sperm and gonadal ducts.

Constructs. Reporter genes were designed using previously published enhancer sequences: *Dmrt.a*³⁰, *Mxsb*^{17,31,32}, *Six1/2*¹⁴, *GnRH*¹⁴, *CNGA*¹⁴, *Foxc*¹² and *islet*⁶. The *Eya* (Ciinte.REG.KhC7:6052317–6053527) and *Pax3/7* (Ciinte.REG. KhC10: 876118–879610) enhancers were isolated via PCR using the following primers (5'-ATGCCTGCAGACTCAATTACCGAATTAATT-3' and 5'-GATCGGATCCATATTTCCATCAGAACTTT-3' for *Eya*, 5'-ATGCCTGCAGGTATGACTGTGTAATCTGC-3' and 5'-GATCGGATCCGTTTGTGGTGTGTTCAG-3' for *Pax3/7*) and cloned into the PstI–BamHI restriction site of the pSpCFP vector.

The following fusion genes were used in the experiments presented in Extended Data Fig. 10. For the *Dmrt.a* ΔMO, the target sequence (Ciinte.REG.KhS544:4,240–5,196) was isolated by PCR using the following primers (5'-ATGCGCATGCTAGTAGGGTGGAGGAAGATG-3' and 5'-GATCGGATCCTTGGTTTAACTCTAAAGC-3') and cloned into the SphI–BamHI restriction site of the pSpCFP vector. To generate the pSPDmrt.a construct, we isolated the coding sequence of *Dmrt.a* (Ciinte.CG.KH.S544.3) with following primers (5'-ATGCGCGGCCGATGCAACCGACAGAGGA-3' and 5'-GATCGAATTCTACTTGTCACTTGAGCATG-3' and cloned it into the NotI–EcoRI restriction site of pSpCFP vector. To generate the *Dmrt.a* ΔMO target sequence > *Dmrt.a* construct, the *Dmrt.a* ΔMO target sequence was inserted into the SphI and BamHI site of the pSPDmrt.a construct. To generate the pSPMxsb MO target sequence CFP, the *Mxsb* MO target sequence (AAATTAAAAATGACAGTAAACGAAT) was tagged by inverse PCR. To generate *Mxsb* > *Mxsb* MO target sequence CFP, the enhancer sequence of *Mxsb* was cloned into the XhoI–NotI site of the pSPMxsb MO target sequence CFP. To generate the pSPMxsb construct, we isolated the coding sequence of *CiMxsb* (Ciinte.CG.KH.C2.957) with the following primers (5'-ATGCGCGGCCGATGACAGTAAACGAATCC-3' and 5'-GCTTGATATCCTATCGACTCTCAGTTGGGT-3'). To generate the pSPMxsb mutant (mut) construct, we replaced the coding region of the *Mxsb* MO target sequence from (ATGACAGTAAACGAAT) to (ATGACGGTGAATGAGT) by inverse PCR (changed nucleotides are underlined). The PCR products were digested with NotI and EcoRV and inserted into the NotI and blunted EcoRI sites of pSpCFP. For *Mxsb* > *Mxsb* mut, the *Mxsb* enhancer DNA was cloned into the XhoI–NotI site of pSPMxsb mut. To generate the pSPFoxc MO target sequence CFP, the *Foxc* MO target sequence (GGTTTGATTCTCTATAATGACAATG) was tagged by inverse PCR. To generate *Foxc* > *Foxc* MO target sequence CFP, the enhancer sequence of *Foxc* was cloned into the XhoI–NotI site of pSPFoxc MO target sequence CFP. To generate the pSPFoxc construct, we isolated the coding sequence of *CiFoxc* (Ciinte.CG.KH.L57.25) with the following primers (5'-ATGCGCGGCCGCTATGACAATGCAATCCG-3' and 5'-GATCGAATTCTCAGTACTTAGTGAATCGT-3'). To generate *Foxc* > *Foxc*, the enhancer sequence of *Foxc* was cloned into the XhoI–NotI site of pSPFoxc.

The following fusion genes were used for the experiments shown in Figs. 2, 3 and Extended Data Figs. 3, 4. The pSPDmrt.a, pSPSix1/2 and pSPFoxc fusion genes were prepared using the coding sequence of *CiDmrt.a* (Ciinte.CG.KH.S544.3), *CiSix1/2* (Ciinte.CG.KH.C3.553) and *CiFoxc* (Ciinte.CG.KH.L57.25). These were amplified with the following primers (5'-ATGCGCGGCCGATGCGCAGCCACCCTGGCG-3' and 5'-GATCGAATTCTTACGATCCCATTTTCGACTG-3' for *Six1/2*, 5'-ATGCGCGGCCGCTATGACAATGCAATCCG-3' and 5'-GATCGAATTCTCAGTACTTAGTGAATCGT-3' for *Foxc*). The PCR products were digested with NotI and EcoRI and inserted into the NotI and EcoRI site of pSpCFP. To generate *Dmrt.a* > *Mxsb*, *Dmrt.a* > *Six1/2* fusion genes, the enhancer region of *Dmrt.a* was inserted into the SphI and BamHI sites of pSPMxsb and pSPSix1/2. The *Dmrt.a* > *Foxc* fusion gene was prepared using the enhancer region of *Dmrt.a* inserted in the SphI and NotI sites of pSPFoxc. To generate *Pax3/7* > *Foxc*, the enhancer region of *Pax3/7* was inserted into the XhoI and NotI sites of pSPFoxc. The minimal enhancer of *Six1/2* was amplified with the following primers (5'-TGCCTGCAGCGAAAACAATGGTTTATCCG-3' and 5'-GATCGGATCCTACATGTACGCGCACTTTAA-3') and cloned into the PstI–BamHI restriction site of a reporter construct containing the pSPFoxAa basal promoter and Kaede³³.

Microinjection of antisense MOs and reporter genes. MOs were obtained from Gene Tools. MOs targeting *Dmrt.a*, *Mxsb*, *Otx* and *Foxc* have previously been described¹⁵. The following MO sequences were used: *Dmrt.a*, 5'-CTGTTTGC TATAATTCTGTAACTC-3'; *Mxsb*, 5'-ATTCGTTTACTGTCAATTTTA ATTT-3'; *Otx*, 5'-TACGACATGTTAGGAATTGAACCCG-3'. *Foxc*

5'-CATTGTTCATTATAGAGAATCAAACC-3'. For control injections we used a universal control MO obtained from Gene Tools. MOs were dissolved in DEPC-treated water containing 1 mg ml⁻¹ tetramethylrhodamine dextran (D1817, Invitrogen). The concentrations of MO and plasmid DNA in the injection medium were 0.5 mM and 1–10 ng μl⁻¹, respectively. Microinjections of MOs and reporter constructs were performed as previously described³⁴. All experiments were repeated at least twice with different batches of embryos. Efficiency and specificity of MOs were evaluated by the simultaneous injection of MO and CFP reporter genes (10 ng μl⁻¹), in which the initiating codon (ATG) was replaced with a nucleotide sequence recognized by individual MOs (Extended Data Fig. 10). The concentration of rescue construct in the injection medium is 1 ng μl⁻¹.

DiI labelling. DiI or DiO labelling of the a5.3, a5.4, b5.3 blastomeres was performed as previously described^{35,36}. DiI (Celltracker CM-DiI Dye, C7000, Molecular Probes) and DiO (D-275, Molecular Probes) were dissolved in soybean oil at a concentration of 5 mg ml⁻¹ (see Extended Data Fig. 1).

Single-cell RNA-seq assays. *Pax3/7* > *Foxc* (2.5 ng μl⁻¹) injected eggs and control eggs were fertilized side by side, and allowed to develop to the late tailbud stage (12 h after fertilization at 18°C). For each sample, 120 morphologically normal embryos were transferred into a 1.5-ml centrifuge tube that was pre-coated with 5% BSA in Ca²⁺-free artificial sea water (Ca²⁺-free ASW, 10 mM KCl, 40 mM MgCl₂, 15 mM MgSO₄, 435 mM NaCl, 2.5 mM NaHCO₃, 7 mM Tris base, 13 mM Tris-HCl). Cells were subsequently dissociated with 300 μl 1% trypsin in Ca²⁺-free ASW with 5 mM EGTA for 5 min. Embryos were pipetted 5 min on ice to complete dissociation of individual cells. Subsequently, 500 μl ice-cold Ca²⁺-free ASW containing 0.5% BSA was added to stop digestion. Cells were collected by centrifuging at 900g for 5 min at 4°C and then resuspended in 50 μl ice-cold Ca²⁺-free ASW containing 0.5% BSA.

Single-cell suspensions were loaded onto the 10X Genomics Chromium system using Reagent Kits to generate and amplify cDNAs, as recommended by the manufacturer (10X Genomics)³⁷. Illumina sequencing libraries were generated from the cDNA samples using the Nextera DNA library prep kit and sequenced using Illumina HiSeq 2500 Rapid flowcells (Illumina) with paired-end 26 + 125 nucleotide reads following standard Illumina protocols. Raw sequencing reads were filtered by Illumina HiSeq Control Software and only pass-filter reads were used for further analysis.

The *Pax3/7* > *Foxc* and wild-type samples were run on both lanes of a HiSeq 2500 Rapid Run mode flow cell. Base calling was performed by Illumina RTA version 1.18.64.0. BCL files were then converted to FASTQ format using bcl2fastq version 1.8.4 (Illumina). Reads that aligned to phix (using Bowtie version 1.1.1) were removed as well as reads that failed Illumina's default chastity filter. We then combined the FASTQ files from each lane and separated the samples using the barcode sequences allowing 1 mismatch (using barcode_splitter version 0.18.2).

Using 10X Cell Ranger version 2.0.1, the count pipeline was run with default settings on the FASTQ files to generate gene–barcode matrices for each sample. The reference sequence was obtained from the Ghost database³⁸ with the *sv40* sequences added. The gene annotations used were also obtained from the Ghost database, again with *sv40* added.

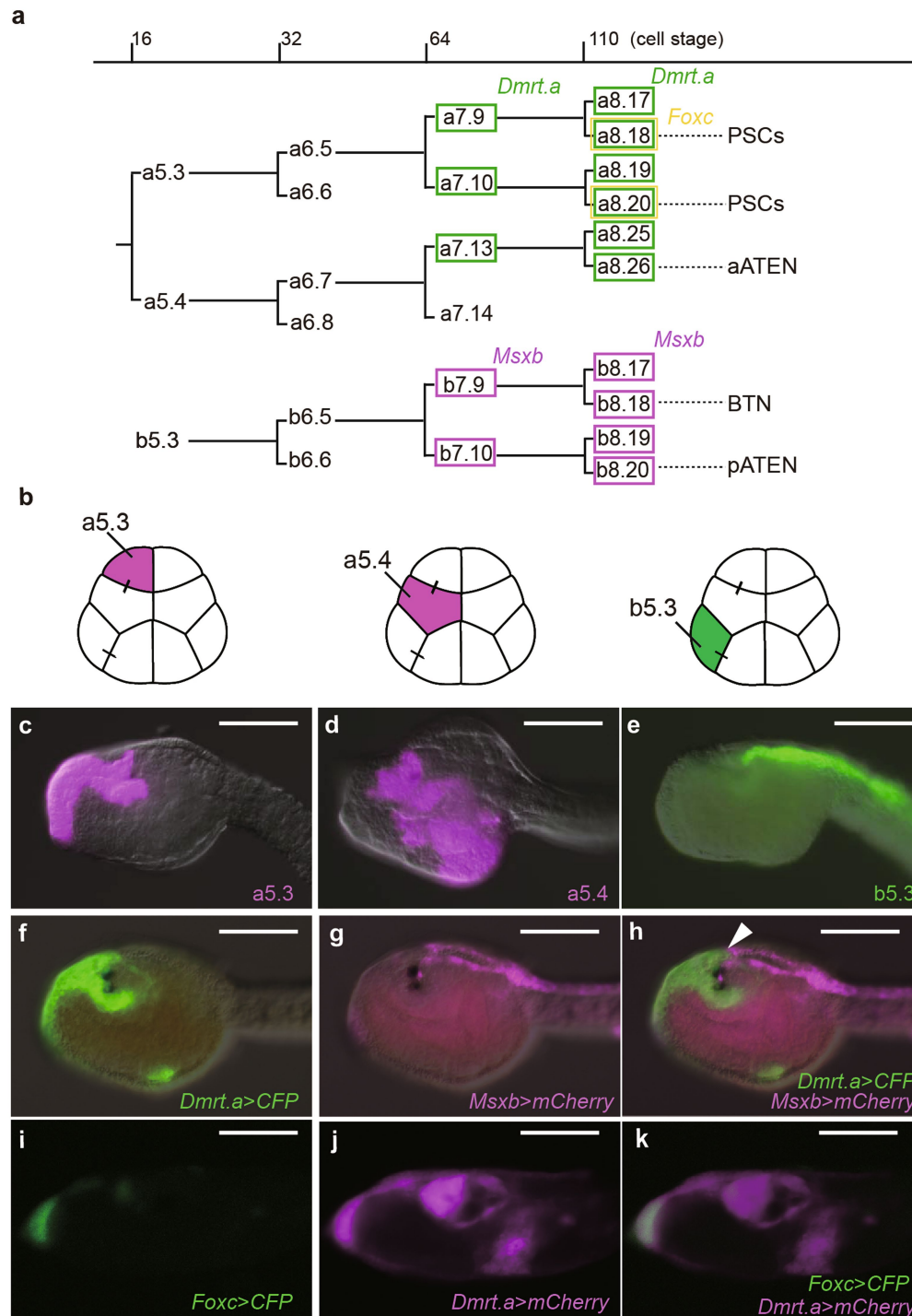
Low-quality transcriptomes were filtered as follows: (1) we discarded cells with less than 200 expressed genes; (2) we discarded cells with less than 500 or more than 30,000 unique molecular identifiers. We further normalized the read counts of each cell by Seurat methods, and the normalized read counts were log₂-transformed for downstream analyses and visualizations. For dimensional reduction, the relative expression measurement of each gene was used to remove unwanted variation. Genes with the top 2,000 highest standard deviations were obtained as highly variable genes of wild-type and transgenic (*Pax3/7* > *Foxc*) samples. We further aligned these two samples using canonical correlation analysis to focus on shared similarities and to facilitate comparative analysis. In the aligned dataset, 10,135 cells were kept and clustered based on their principal component analysis scores with highly variable genes. Basically, after significant principal components were identified, a graph-based clustering approach was used for partitioning the cellular distance matrix into clusters. Cell distance was visualized by *t*-SNE in reduced 2D space. Differentially expressed genes between different cell types were identified by the following criteria using the DESeq2 software package: (1) false-discovery rate-adjusted *P* < 0.01; (2) absolute log₂(fold change) between groups were larger than 1.

Image acquisition. Images of transgenic larvae were obtained with a Zeiss AX 10 epifluorescence microscope.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Single-cell RNA-seq data that support the findings of this study have been deposited in Gene Expression Omnibus (GEO) with the accession code GSE115331. All other data that support the findings of this study are available from corresponding authors upon reasonable request.

30. Gainous, T. B., Wagner, E. & Levine, M. Diverse ETS transcription factors mediate FGF signaling in the *Ciona* anterior neural plate. *Dev. Biol.* **399**, 218–225 (2015).
31. Russo, M. T. et al. Regulatory elements controlling *Ci-msxb* tissue-specific expression during *Ciona intestinalis* embryonic development. *Dev. Biol.* **267**, 517–528 (2004).
32. Roure, A., Lemaire, P. & Darras, S. An *otx/nodal* regulatory signature for posterior neural development in ascidians. *PLoS Genet.* **10**, e1004548 (2014).
33. Hozumi, A. et al. Enhancer activity sensitive to the orientation of the gene it regulates in the chordate genome. *Dev. Biol.* **375**, 79–91 (2013).
34. Satou, Y., Imai, K. S. & Satoh, N. Action of morpholinos in *Ciona* embryos. *Genesis* **30**, 103–106 (2001).
35. Satou, Y., Imai, K. S. & Satoh, N. The ascidian *Mesp* gene specifies heart precursor cells. *Development* **131**, 2533–2541 (2004).
36. Horie, T. et al. Ependymal cells of chordate larvae are stem-like cells that form the adult nervous system. *Nature* **469**, 525–528 (2011).
37. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
38. Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. & Satoh, N. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* **22**, 837–843 (2005).



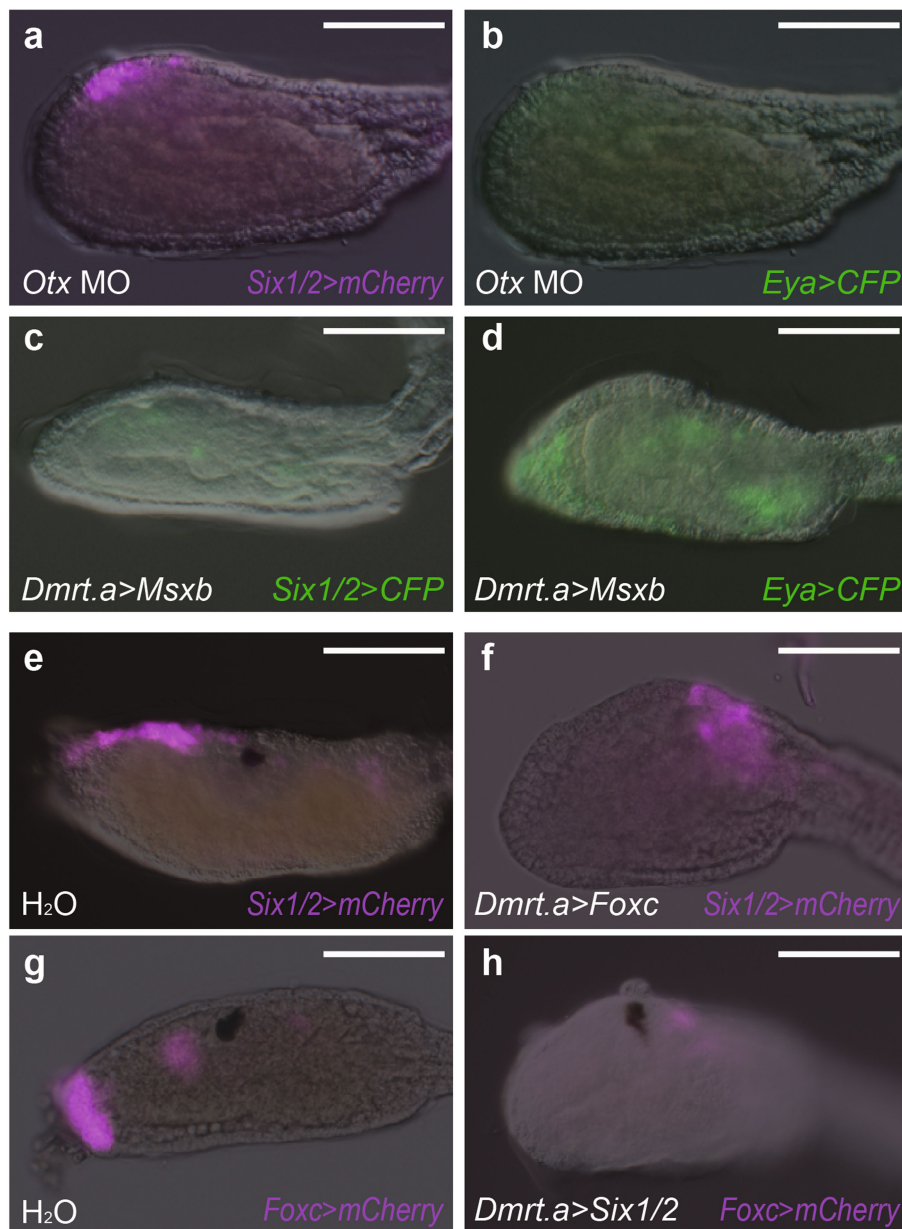
Extended Data Fig. 1 | Sensory cell lineages. a, Cell lineages of anterior blastomeres (a-line blastomeres) and a posterior blastomere (b-line) from the 16-cell to 110-cell stages. Green, *Dmrt.a*-expressing blastomeres; magenta, *Msxb* expression lineages; yellow, *Foxc* expression. **b**, Schematic of 16-cell-stage embryos. Each of the blastomeres that was labelled with DiI or DiO is indicated by magenta or green, respectively. **c–e**, Head regions of larvae labelled with DiI or DiO at the 16-cell stage.

c, Labelling of the a5.3 lineage. **d**, Labelling of the a5.4 lineage. **e**, Labelling of the b5.3 blastomere. **f–h**, Head region of a larva that was injected with *Dmrt.a* > CFP and *Msxb* > mCherry reporter genes. Arrowhead identifies the boundary of the *Dmrt.a*–*Msxb* expression territories. **i–k**, Head region of a larva injected with *Foxc* > CFP and *Dmrt.a* > mCherry reporter genes. Anterior is to the left. Scale bars, 100 μ m.



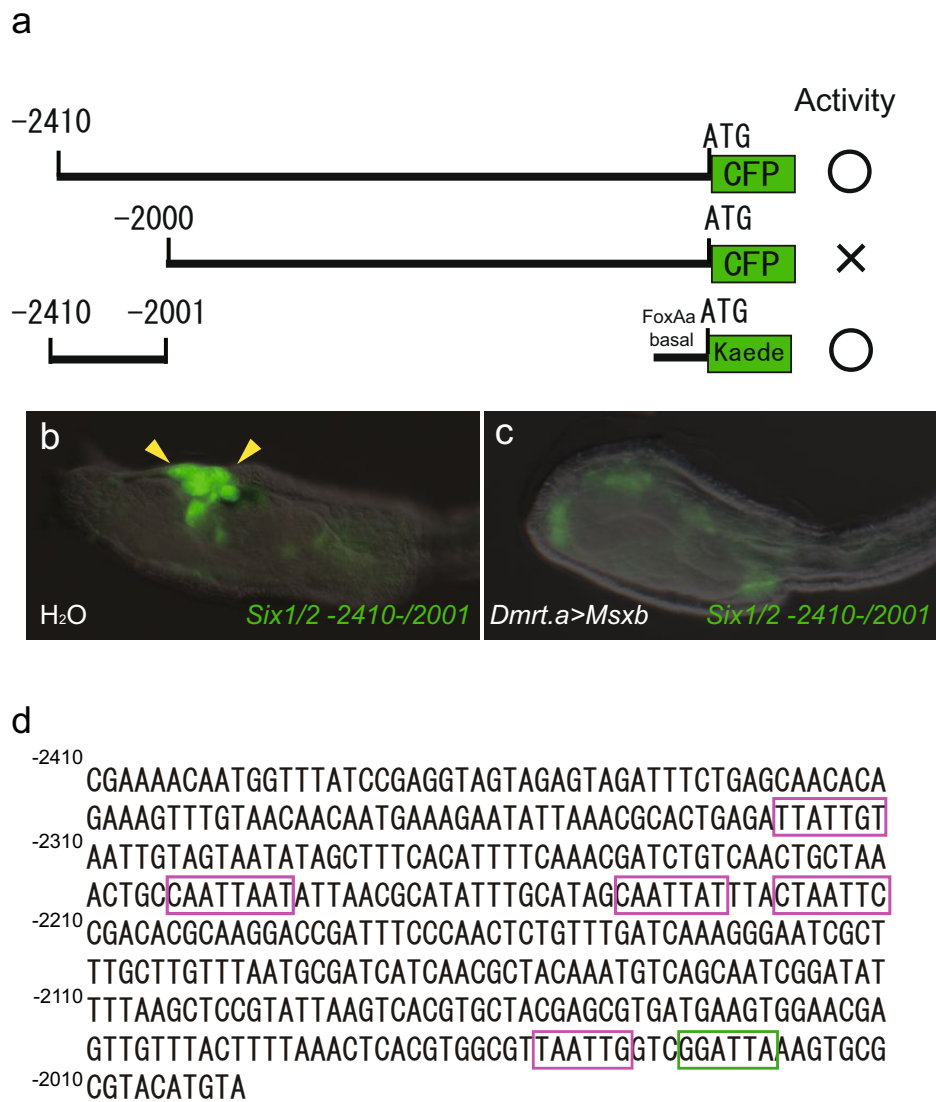
Extended Data Fig. 2 | Regulation of *Eya* expression by *Dmrt.a* and *Msxb*. **a–c**, Head regions of larvae injected with an *Eya* > *CFP* reporter gene. Yellow arrowheads denote *Eya* expression in the proto-placodal region of control MO-injected larvae (36 out of 36 larvae displayed this

pattern) (**a**). **b**, There is a loss of expression in *Dmrt.a* morphants (88 out of 88 larvae showed this phenotype). **c**, There is expanded expression (white arrowheads) in *Msxb* morphants (39 out of 48 larvae showed this phenotype). Anterior is to the left. Scale bars, 100 μm.



Extended Data Fig. 3 | Regulatory interactions among placodal determinants. **a, b,** Head regions of larvae that were injected with an *Otx* MO, and injected with *Six1/2 > mCherry* (**a**) and *Eya > CFP* (**b**) (32 out of 32 larvae showed reduced or no expression of *Six1/2 > mCherry* and no expression of *Eya > CFP*). **c, d,** Head regions of larvae injected with *Dmrt.a > Msxb*, and injected with *Six1/2 > CFP* (**c**; 33 out of 33 larvae showed no expression of *Six1/2 > CFP*) and *Eya > CFP* (**d**; 71 out of 71 larvae showed little or no expression of *Eya > CFP*). **e, f,** Head regions of

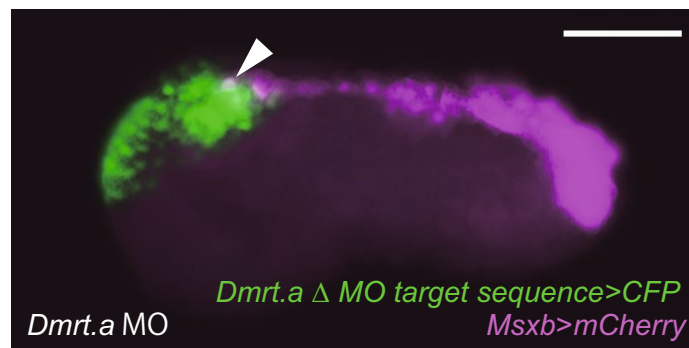
larvae injected with *Six1/2 > mCherry*, and injected with *H₂O* (**e**; 61 out of 61 larvae showed full expression of *Six1/2 > mCherry*) or *Dmrt.a > Foxc* (**f**; 50 out of 50 larvae showed no expression of *Six1/2 > mCherry*). **g, h,** Head regions of larvae injected with *Foxc > mCherry*, and injected with *H₂O* (**g**; 40 out of 40 larvae showed full expression of *Foxc > mCherry*) or *Dmrt.a > Six1/2* (**h**; 39 out of 39 larvae showed no expression of *Foxc > mCherry*). Scale bars, 100 μm.



Extended Data Fig. 4 | Direct repression of *Six1/2* expression by *Msxb*.

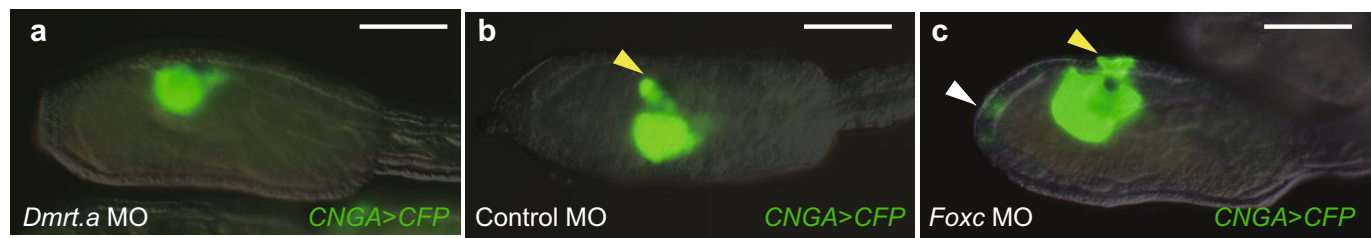
a, Deletion analyses of the 5' regulatory region of *Six1/2*. -2,410 bp to -2,001 bp of the 5' cis-regulatory region is necessary for *Kaede* expression in the pre-placodal territory. **b**, **c**, Head regions of larvae injected with *Six1/2* -2,410 to -2,001 > *Kaede*, and injected with H₂O (**b**; 74 out of

74 larvae showed full expression of *Six1/2* -2,410 to -2,001 > *Kaede*) or *Dmrt.a* > *Msxb* (**c**; 37 out of 37 larvae showed no expression of *Six1/2* -2,410 to -2,001 > *Kaede*). **d**, The *Six1/2* 5' regulatory region spanning -2,410 to -2,001 bp contains an *Otx* binding site (green box) and multiple *Msxb* repressor binding sites (magenta boxes).



Extended Data Fig. 5 | Anterior expansion of *Msxb* expression in *Dmrt.a* morphant. Tailbud embryo injected with *Dmrt.a* MO, *Dmrt.a* Δ MO target sequence > CFP and *Msxb* > mCherry construct. The anterior

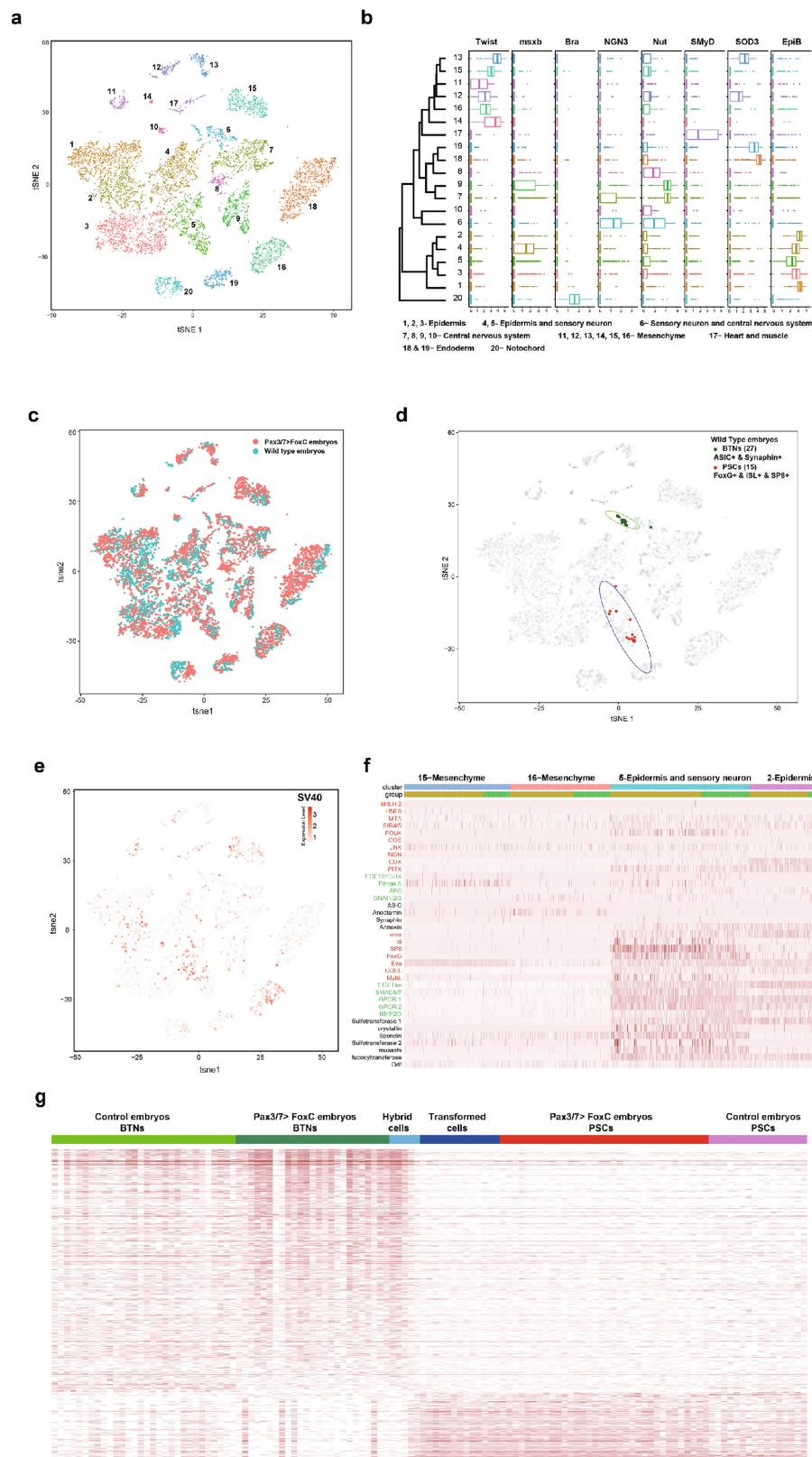
expansion of the *Msxb* > mCherry expression pattern is indicated by the white arrowhead. Anterior is to the left. Scale bars, 100 μ m.



Extended Data Fig. 6 | Specification of aATEN sensory neurons.

a–c, Larvae injected with the *CNGA > CFP* reporter gene. Yellow arrowheads indicate expression in aATENs, and white arrowheads indicate ectopic sites of differentiated aATENs in the palp. **a**, Larva injected with *Dmrt.a* MO (62 out of 62 larvae showed no expression of *CNGA > CFP* in

aATENs). **b**, Larva injected with control MO (32 out of 32 larvae showed full expression of *CNGA > CFP*). **c**, Larva injected with *Foxc* MO (18 out of 42 larvae showed expanded expression of *CNGA > CFP* in palp region). Anterior is to the left. Scale bars, 100 μm.

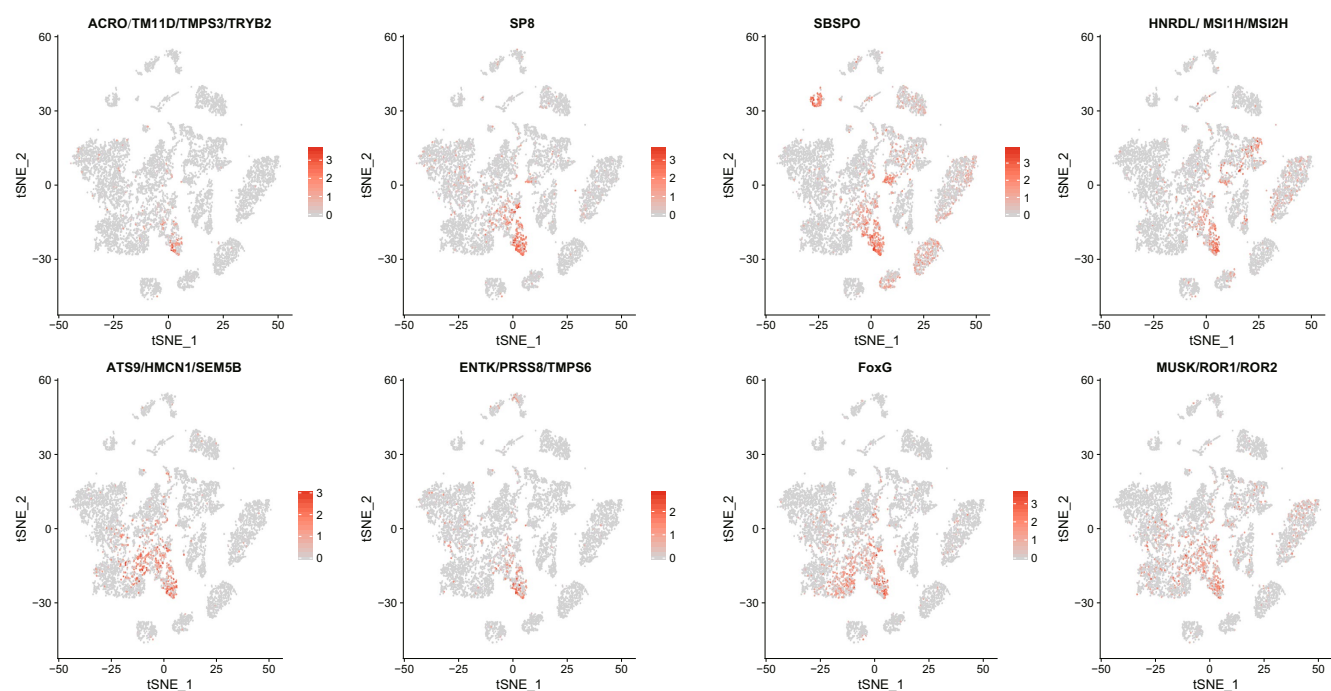


Extended Data Fig. 7 | See next page for caption.

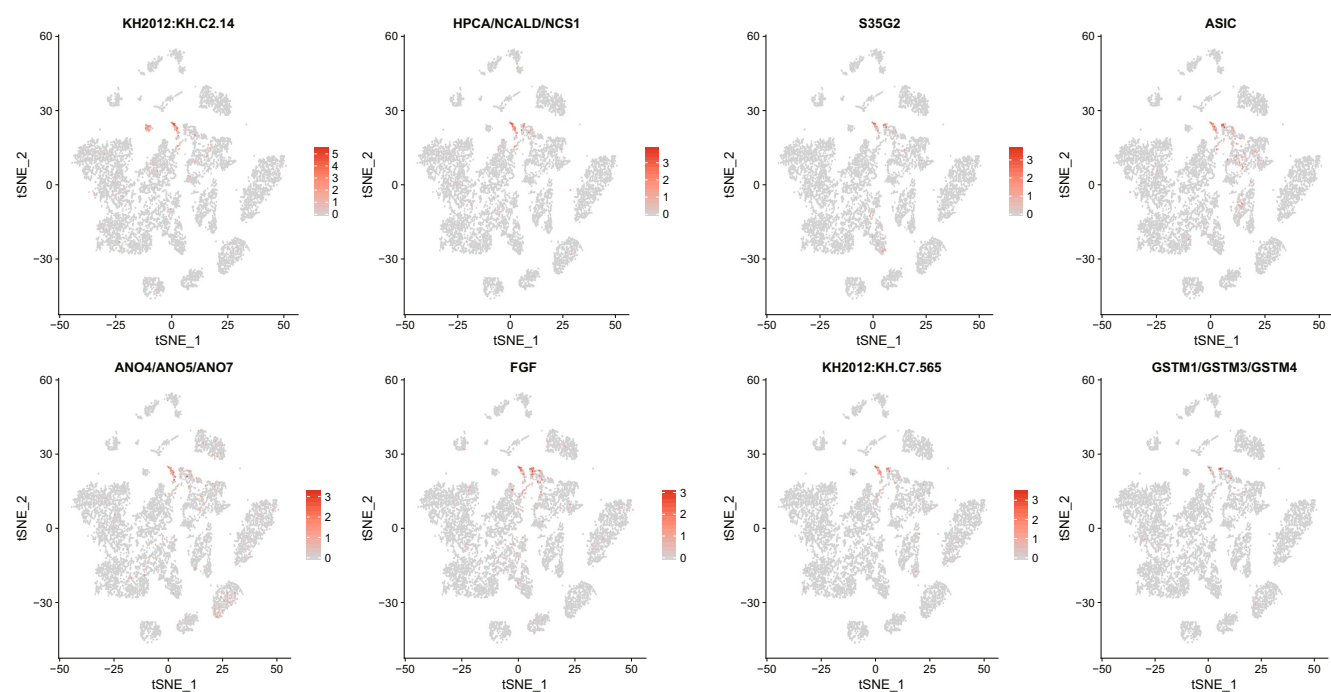
Extended Data Fig. 7 | Single-cell RNA-seq analysis. a, b, Identification of major cell types among 20 clusters with known markers, 3 epidermis clusters were identified by the expression of *EpiB*; 2 clusters of epidermis and sensory neurons were identified by the expression of either *Msx* or *nut*, along with *EpiB*; a cluster of sensory neurons identified by the expression of *NGN3* and *nut*; 3 clusters of central nervous system by high levels of *nut* transcripts; 6 clusters of mesenchyme by their expression of *twist*; 2 endoderm clusters by the high expression of *SOD3*; and one notochord cluster by the expression of *Ci* (*T* also known as *brachyury*). **c,** Representation of overlap between cells from *Pax3/7 > Foxc* and control embryos in each cell population. *t*-SNE plot from Fig. 3a, with each cell now coloured to indicate their origin from either *Pax3/7 > Foxc* embryos (red dots, $n = 5,339$) or control embryos (blue dots, $n = 4,850$). Both samples contribute to all 20 cell populations. **d,** Identification of BTNs

and PSCs with the combination of representation markers in the control embryo, 27 BTNs (green dots) were identified by the combination of *Asic1b* and *synaphin*, and 15 PSCs were identified by the combination of *Foxg*, *islet* and *SP8*. **e,** Visualization of SV40⁺ cells in *Pax3/7 > Foxc* transgenic embryos within *t*-SNE projection map. SV40 is detected in cells contained within clusters 2 and 5 for epidermis and sensory neuron cells, as well as weak expression in mesenchyme clusters 15 and 16. None of the transformed or hybrid cells contained in the sensory cell clusters (5 and 6) express any mesenchyme marker genes, suggesting that none of these are transformed by misexpression of *Pax3/7*. **f,** Heat map of representative genes (Fig. 3e) that show no significant differential expression in SV40⁺ cells contained within clusters 2, 5, 15 and 16. **g,** Heat map of all differentially expressed genes between BTNs and PSCs from both control and *Pax3/7 > Foxc* embryos.

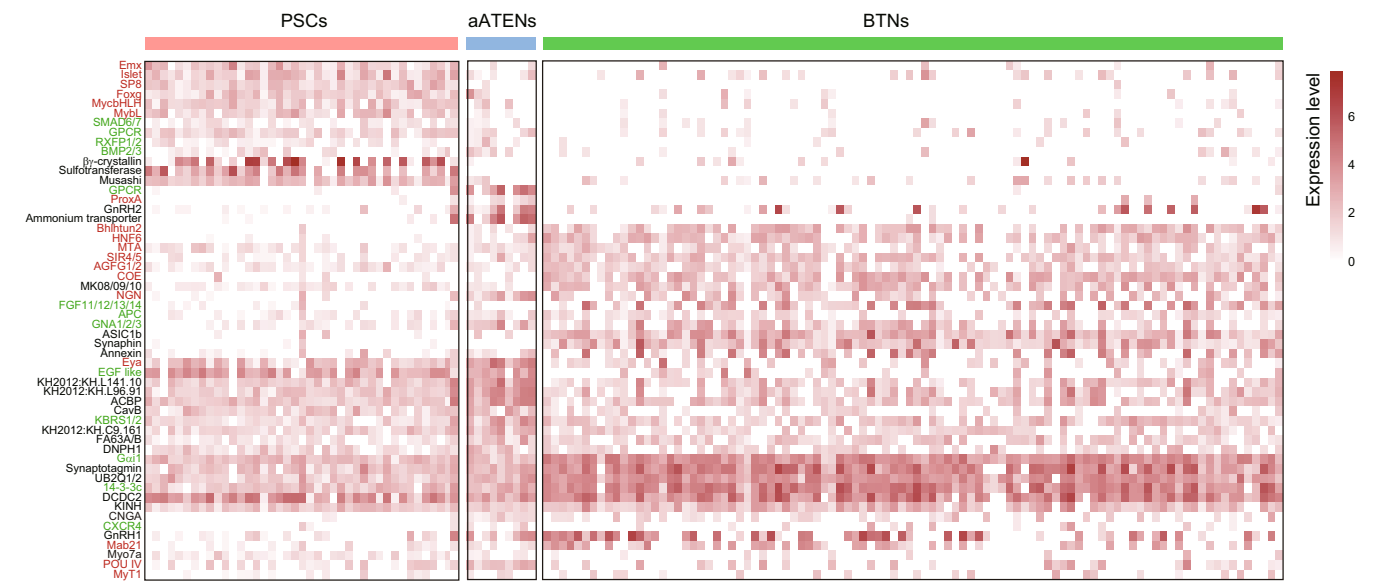
a



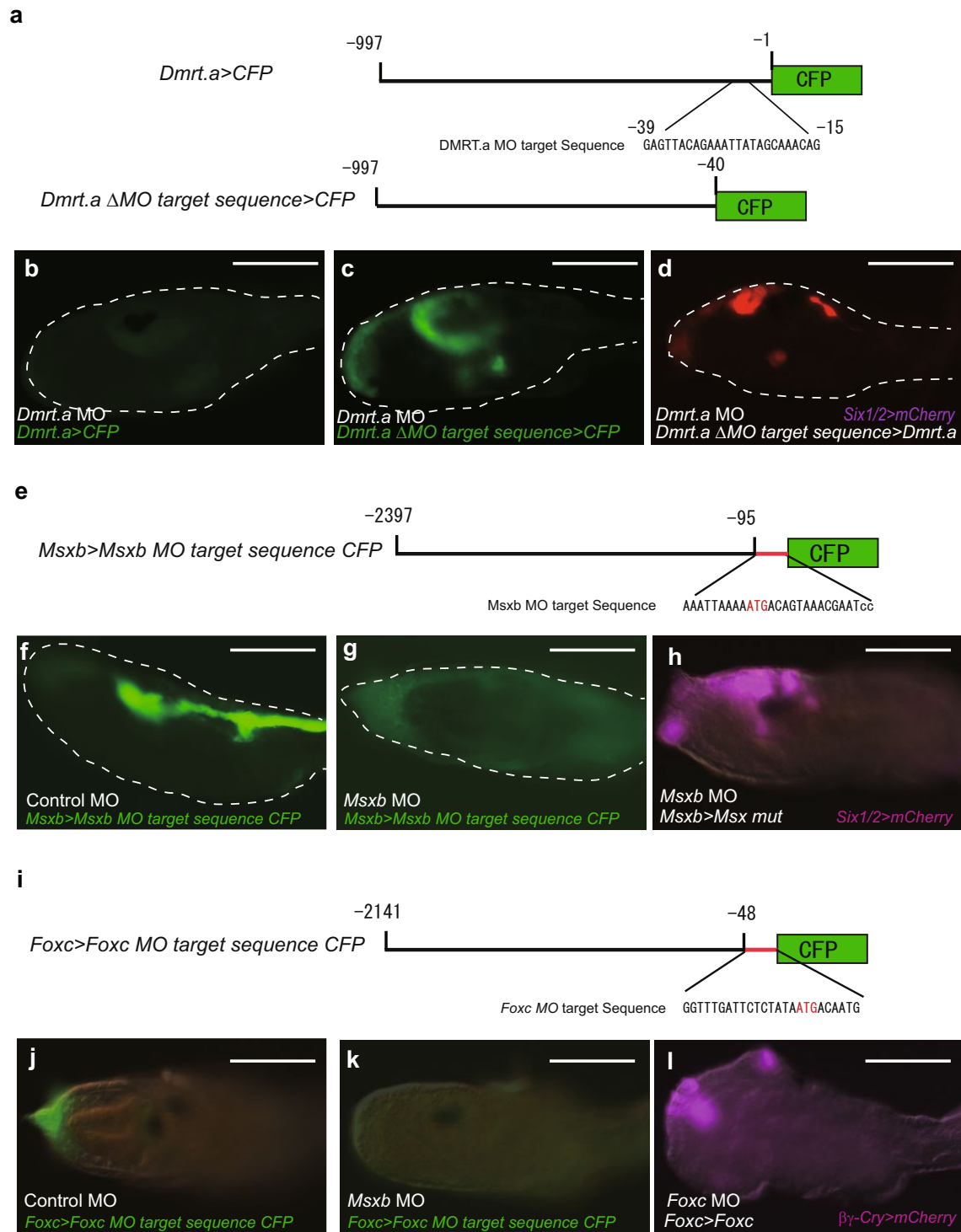
b



Extended Data Fig. 8 | Newly identified markers for PSCs and BTNs. Distribution of newly identified marker genes in PSCs (a) and BTNs (b).



Extended Data Fig. 9 | Heat map of differentially expressed and coexpressed genes between PSCs, aATENs and BTNs from wild-type late tailbud stage II embryos. Transcription factors (red), signalling pathway genes (green) and effector genes (black).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Control experiments for MO gene disruption assays. a–d, *Dmrt.a* MO. **a**, Schematic of reporter gene containing *Dmrt.a* regulatory genes with and without recognition sequences for the *Dmrt.a* MO that was used in this study. The MO recognition sequences are located in the 5' UTR, upstream of the initiating AUG codon (–1). **b**, Larva injected with *Dmrt.a* MO and injected with the *Dmrt.a* > CFP reporter gene containing the MO recognition sequences (46 out of 46 larvae showed no protein synthesis from the *Dmrt.a* > CFP reporter). **c**, Same as **b**, except that the reporter gene lacks the *Dmrt.a* MO recognition sequence (*Dmrt.a* ΔMO target > CFP) (57 out of 57 larvae showed CFP expression in appropriate head tissues). *Dmrt.a* MO efficiently blocks the expression of CFP that contains the *Dmrt.a* MO target sequence. **d**, Larva injected with *Dmrt.a* MO, *Dmrt.a* ΔMO > *Dmrt.a* and *Six1/2* > *mCherry*. *Dmrt.a* MO morphants normally lack *Six1/2* > *mCherry* expression (Fig. 2b), but expression is restored with a *Dmrt.a* transgene that lacks the MO recognition sequence (107 out of 108 larvae showed expression of *Six1/2* > *mCherry*). This result shows that the *Dmrt.a* MO used in this study specifically blocks the synthesis of *Dmrt.a* protein products. **e–h, *Msx*b MO.** **e**, Diagram of *Msx*b 5' regulatory region and the location of the recognition sequences for the *Msx*b MO and point mutations in this sequence. **f, g**, Larvae injected with *Msx*b > CFP containing MO recognition sequences, and injected with control MO (**f**; 54 out of 54 larvae showed CFP expression). **g**, Same as **f** except that the *Msx*b

MO was injected instead of the control MO (44 out of 44 larvae showed no expression of CFP). These results show that the *Msx*b MO specifically blocks CFP protein synthesis from the *Msx*b reporter gene. **h**, Larva injected with *Msx*b MO, *Six1/2* > *mCherry* and *Msx*b > CFP reporter gene containing point mutations in MO recognition sequence (see red letters in **e**). *Msx*b morphants normally display expanded expression of *Six1/2* in tail regions (Fig. 2c). This expansion is suppressed by injection of the mutant *Msx*b transgene lacking the MO recognition sequences (**h**). This result suggests that the *Msx*b MO inhibits synthesis of *Msx*b protein products. **i–l, *Foxc* MO.** **i**, Diagram of *Foxc* 5' regulatory region showing the location of the MO recognition sequence and point mutations in this sequence. **j–l**, Larvae injected with *Foxc* > *Foxc* transgene and *Foxc* > CFP reporter gene, and also injected with control MO (**j**; 25 out of 25 larvae showed CFP expression). **k**, Same as **j** except that the embryo was injected with the *Foxc* MO instead of the control MO (**k**; 99 out of 99 larvae showed no expression of CFP). This result shows that the *Foxc* MO efficiently blocks the synthesis of CFP proteins encoded by the *Foxc* > CFP reporter gene containing the *Foxc* MO target sequence. **l**, Larvae injected with *Foxc* MO, *Foxc* > *Foxc* mut (MO-resistant *Foxc* cDNA) and $\beta\gamma$ -crystallin > *mCherry*. Normally, *Foxc* morphants lack expression of the $\beta\gamma$ -crystallin > *mCherry* reporter gene (Fig. 2f). However, expression is restored by injection of *Foxc* > *Foxc* transgene. This result suggests that the *Foxc* MO inhibits synthesis of *Foxc* protein products. Scale bars, 100 μ m.

Structure and function of the global topsoil microbiome

Mohammad Bahram^{1,2,3,21*}, Falk Hildebrand^{4,21}, Sofia K. Forslund^{4,16,17}, Jennifer L. Anderson², Nadejda A. Soudzilovskaia⁵, Peter M. Bodegom⁵, Johan Bengtsson-Palme^{6,7,18}, Sten Anslan^{1,8}, Luis Pedro Coelho⁴, Helery Harend¹, Jaime Huerta-Cepas^{4,19}, Marnix H. Medema⁹, Mia R. Maltz¹⁰, Sunil Mundra¹¹, Pål Axel Olsson¹², Mari Pent¹, Sergei Pölme¹, Shinichi Sunagawa^{4,20}, Martin Ryberg², Leho Tedersoo^{13*} & Peer Bork^{4,14,15*}

Soils harbour some of the most diverse microbiomes on Earth and are essential for both nutrient cycling and carbon storage. To understand soil functioning, it is necessary to model the global distribution patterns and functional gene repertoires of soil microorganisms, as well as the biotic and environmental associations between the diversity and structure of both bacterial and fungal soil communities^{1–4}. Here we show, by leveraging metagenomics and metabarcoding of global topsoil samples (189 sites, 7,560 subsamples), that bacterial, but not fungal, genetic diversity is highest in temperate habitats and that microbial gene composition varies more strongly with environmental variables than with geographic distance. We demonstrate that fungi and bacteria show global niche differentiation that is associated with contrasting diversity responses to precipitation and soil pH. Furthermore, we provide evidence for strong bacterial–fungal antagonism, inferred from antibiotic-resistance genes, in topsoil and ocean habitats, indicating the substantial role of biotic interactions in shaping microbial communities. Our results suggest that both competition and environmental filtering affect the abundance, composition and encoded gene functions of bacterial and fungal communities, indicating that the relative contributions of these microorganisms to global nutrient cycling varies spatially.

Bacteria and fungi dominate terrestrial soil habitats in terms of biodiversity, biomass and their influence over essential soil processes⁵. Specific roles of microbial communities in biogeochemical processes are reflected by their taxonomic composition, biotic interactions and gene functional potential^{1–4}. Although microbial-biogeography studies have focused largely on single taxonomic groups, and on how their diversity and composition respond to local abiotic soil factors (for example, pH^{6,7}), global patterns and the impact of biotic interactions on microbial biogeography remain relatively unexplored. In addition to constraints imposed by environmental factors, biotic interactions may strongly influence bacterial communities. For example, to outcompete bacteria, many fungal taxa secrete substantial amounts of antimicrobial compounds⁸, which may select for antibiotic-resistant bacteria and effectively increase the relative abundance of antibiotic-resistance genes (ARGs). Here we used metagenomics and DNA metabarcoding (16S, 18S and internal transcribed spacer (ITS) rRNA gene markers), soil chemistry and biomass assessments (phospholipid fatty acids analyses (PLFAs)) to determine the relationships among genetic (functional

potential), phylogenetic and taxonomic diversity and abundance in response to biotic and abiotic factors in 189 topsoil samples, covering all terrestrial regions and biomes of the world⁹ (Extended Data Fig. 1a, Supplementary Table 1). Altogether, 58,000 topsoil subsamples were collected from 0.25-ha plots from 1,450 sites (40 subsamples per site), harbouring homogeneous vegetation that were minimally affected by humans. We minimized biases and shortcomings in sampling¹⁰ as well as technical variation, including batch effects¹¹, by using highly standardized collection and processing protocols. From the total collection, 189 representative sites were selected for this analysis. We validated our main findings in external datasets, including an independent soil dataset (145 topsoil samples; Supplementary Table 1) that followed the same sampling and sequencing protocol.

Using metagenomics, we constructed a gene catalogue for soils, by combining our newly generated data with published soil metagenomes ($n = 859$, Supplementary Table 1) and identified 159,907,547 unique genes (or fragments thereof). Only 0.51% of these 160 million genes overlapped with those from published genomes and large gut¹² and ocean¹³ gene catalogues that are much closer to saturation (Supplementary Table 2), indicating that the functional potential of soil microbiomes is enormously vast and undersampled. For functional analysis, we annotated genes and functional modules via orthologous groups using the eggNOG database¹⁴. For each sample, we also constructed taxonomic profiles at the class and phylum levels for both bacteria and fungi from relative abundance of rRNA genes in metagenomic datasets (miTags¹⁵), complemented by operational taxonomic units (OTUs) that were based on clustering 18S rRNA and ITS¹⁶ genes for soil fungi and 16S rRNA genes for soil bacteria at 97% similarity threshold (see Methods ‘Metagenomics and metabarcoding analyses’). In total, 34,522 16S-based bacterial, 2,086 18S-based and 33,476 ITS-based fungal OTUs were analysed in the context of geographic space and 16 edaphic and climatic parameters were determined for each sampling site (see Methods ‘Statistical analyses’). Archaea were poorly represented in our metabarcoding data (less than 1% of OTUs) and metagenomics data (less than 1% of miTags) and hence are excluded from most analyses.

We examined whether the latitudinal diversity gradient (LDG), a trend of increasing diversity from the poles to the tropics seen in many macroscopic organisms, especially plants¹⁷, applies to microbial global distribution patterns¹⁰. We found that, contrary to the typical LDG,

¹Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia. ²Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. ³Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden. ⁴Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. ⁵Environmental Biology Department, Institute of Environmental Sciences, CML, Leiden University, Leiden, The Netherlands. ⁶Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Göteborg, Göteborg, Sweden. ⁷Centre for Antibiotic Resistance research (CARE), University of Göteborg, Göteborg, Sweden. ⁸Braunschweig University of Technology, Zoological Institute, Braunschweig, Germany. ⁹Bioinformatics Group, Wageningen University, Wageningen, The Netherlands. ¹⁰Center for Conservation Biology, University of California, Riverside, Riverside, CA, USA. ¹¹Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, Oslo, Norway. ¹²Biodiversity Unit, Department of Biology, Ecology building, Lund University, Lund, Sweden. ¹³Natural History Museum, University of Tartu, Tartu, Estonia. ¹⁴Max Delbrück Centre for Molecular Medicine, Berlin, Germany. ¹⁵Department of Bioinformatics, University of Würzburg, Würzburg, Germany. ¹⁶Present address: Experimental and Clinical Research Center, a cooperation of Charité-Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany. ¹⁷Present address: Max Delbrück Centre for Molecular Medicine, Berlin, Germany. ¹⁸Present address: Wisconsin Institute of Discovery, University of Wisconsin-Madison, Madison, WI, USA. ¹⁹Present address: Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain. ²⁰Present address: Department of Biology, Institute of Microbiology, ETH Zurich, Zurich, Switzerland. ²¹These authors contributed equally: Mohammad Bahram, Falk Hildebrand. *e-mail: bahram@ut.ee; leho.tedersoo@ut.ee; bork@embl.de

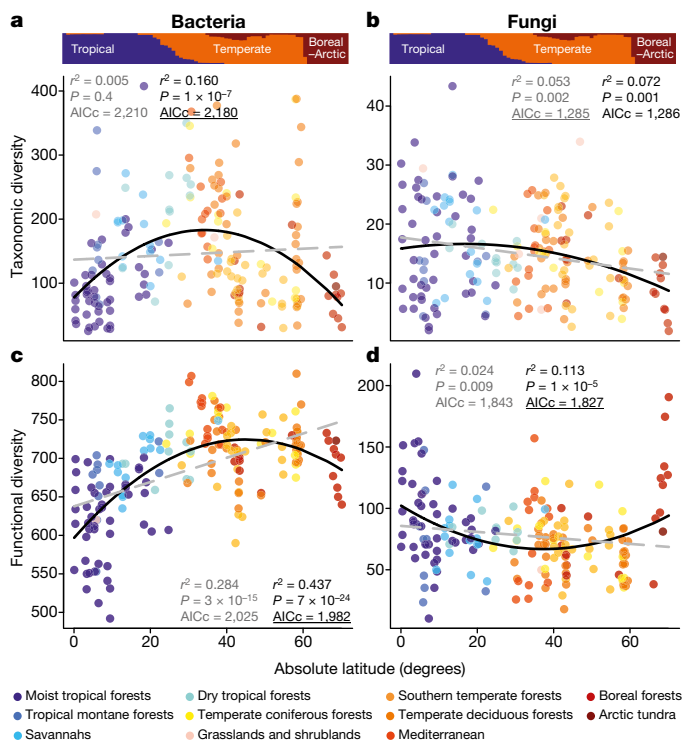


Fig. 1 | Fungal and bacterial diversity exhibit contrasting patterns across the latitudinal gradient. a–d. Latitudinal distributions of bacterial (a, c) and fungal (b, d) taxonomic (a, b; $n = 188$ biologically independent samples) and gene functional (c, d; $n = 189$ biologically independent samples) diversity in global soil samples. First- and second-order polynomial fits are shown in grey and black, respectively. The best polynomial fit was determined (as underlined) on the basis of the corrected Akaike Information Criterion (AICc; see Methods ‘Statistical analyses’) of the first and second order polynomial models (ANOVA: a, $F = 34.28$, $P < 10^{-7}$; b, $F = 3.84$, $P = 0.052$; c, $F = 50.48$, $P < 10^{-10}$; d, $F = 18.55$, $P < 10^{-4}$). Grey dashed and black solid lines are the first and second order polynomial regression lines, respectively. Diversity was measured using inverse Simpson index (these trends were robust to the choice of index, see Extended Data Fig. 2b, c). The latitudinal distribution of the high-level biome (tropical, temperate and boreal-arctic) is given at the top of a and b.

both taxonomic and gene functional diversity of bacteria peaked at mid-latitudes and declined towards the poles and the equator, as is also seen in the global ocean¹³, although the pattern was relatively weak for taxonomic diversity herein (Fig. 1a, c, Extended Data Figs. 1b, 2). The deviation of several bacterial phyla (5 out of 20) from the general trends may be explained by responses to edaphic and climate factors weakly related to latitude (Extended Data Fig. 1b) or contrasting effects at lower taxonomic levels (Supplementary Discussion). By contrast, the LDG does apply to overall fungal taxonomic diversity, and to three out of five fungal phyla when examined separately, but not to fungal functional diversity, which was lowest in temperate biomes and exhibited an inverse unimodal relationship with latitude (Fig. 1b, d, Extended Data Fig. 2c). The LDG was negligible for oceanic fungi (regression analysis, $P > 0.05$)¹³, possibly owing to their lower dispersal limitation and the paucity of plant associations. Although fungal taxonomic diversity decreased poleward, the total fungal biomass (inferred from PLFA markers) and the fungal/bacterial biomass ratio increased poleward, partly owing to a decline in bacterial biomass with increasing latitude (Extended Data Fig. 3a–c).

We tested the extent to which deterministic processes (such as competition and environmental filtering; that is, the niche theory) versus neutral processes (dispersal and drift; the neutral theory) explain the distributions of fungal and bacterial taxa and functions¹⁸. In bacteria, environmental variation correlated strongly with taxonomic

composition (partial Mantel test accounting for geographic distance between samples: $r_{\text{Env|Geo}} = 0.729$, $P = 0.001$) and moderately with gene functional composition ($r_{\text{Env|Geo}} = 0.100$, $P = 0.001$), whereas the overall effect of geographic distance among samples was negligible ($P > 0.05$). The weak correlation between geographic and taxonomic as well as functional composition suggests that environmental variables are more important than dispersal capacity in determining global distributions of soil bacteria and their encoded functions, as previously suggested¹⁹ and observed for oceanic prokaryotes¹³.

For fungi, both geographic distance and environmental parameters were correlated with taxonomic composition (ITS data: $r_{\text{Geo|Env}} = 0.307$, $P = 0.001$; $r_{\text{Env|Geo}} = 0.208$, $P = 0.001$; 18S data: $r_{\text{Geo|Env}} = 0.193$, $P = 0.001$; $r_{\text{Env|Geo}} = 0.333$, $P = 0.001$). Environmental distance (but not geographic distance) correlated with composition of fungal functional genes ($r_{\text{Env|Geo}} = 0.197$, $P = 0.001$), as was also observed for bacteria. The relatively weaker correlation of fungi with environmental variation is consistent with results from local scales⁷. Thus, at both global and local scales, different processes appear to underlie community assembly of fungi and bacteria.

To more specifically investigate the association between environmental parameters and the distribution of taxa and gene functions on a global scale, we used multiple regression modelling (see Methods ‘Statistical analyses’). We found that bacterial taxonomic diversity, composition, richness and biomass as well as the relative abundance of major bacterial phyla can be explained by soil pH, nutrient concentration and to a lesser extent by climatic variables (Extended Data Figs. 4, 5, Supplementary Table 4). The composition of bacterial communities responded most strongly to soil pH, followed by climatic variables, particularly mean annual precipitation (MAP; Extended Data Figs. 4, 5). This predominant role of pH agrees with studies from local to continental scales⁶, and may be ascribed to the direct effect of pH or related variables such as the concentration of calcium and other cations⁶. The relative abundance of genes that encode several metabolic and transport pathways were strongly increased with pH (Extended Data Fig. 4c), suggesting that there may be greater metabolic demand for these functions for bacteria in high-nutrient and alkaline conditions.

Compared to temperate biomes, tropical and boreal habitats contained more closely related taxa at the tip of phylogenetic trees, but from more distantly related clades (Extended Data Fig. 2d), indicating a deeper evolutionary niche specialization in bacteria²⁰. Together with global biomass patterns (Extended Data Fig. 2a), these results suggest that soil bacterial communities in the tropics and at high latitudes are subjected to stronger environmental filtering and include a relatively greater proportion of edaphic-niche specialists, possibly rendering these communities more vulnerable to global change. By contrast, phylogenetic overdispersion in temperate bacterial communities, may result from greater competitive pressure²⁰ or nutrient availability as predicted by the niche theory²¹.

In contrast to the strong association between bacterial taxonomic diversity and soil pH, diversity of bacterial gene functions was more strongly correlated with MAP (Extended Data Fig. 5a–h). The steeper LDG in gene functions than in taxa (Fig. 1a, c) may thus relate to the stronger association of specific metabolic functions to climate than to local soil conditions. Although soil and climate variables exhibited comparable correlations with fungal taxa, the soil carbon-to-nitrogen ratio (C/N) was the major predictor for fungal biomass and relative abundance and composition of gene functions (Extended Data Figs. 3g, 4b, d, Supplementary Table 4). We hypothesize that, compared to bacteria, the global distribution of fungi is more limited by resource availability owing to specialization for the use of specific compounds as substrates and greater energy demand.

We interpret opposing biogeographic trends for bacteria and fungi as niche segregation, driven by differential responses of bacteria and fungi to environmental factors⁷ and their direct competition. Gene functional diversity of both bacteria and fungi responded to MAP and soil pH, albeit in opposite directions (Extended Data Fig. 5c, d, g, h, Supplementary Table 3). This may partly explain the observed inverse

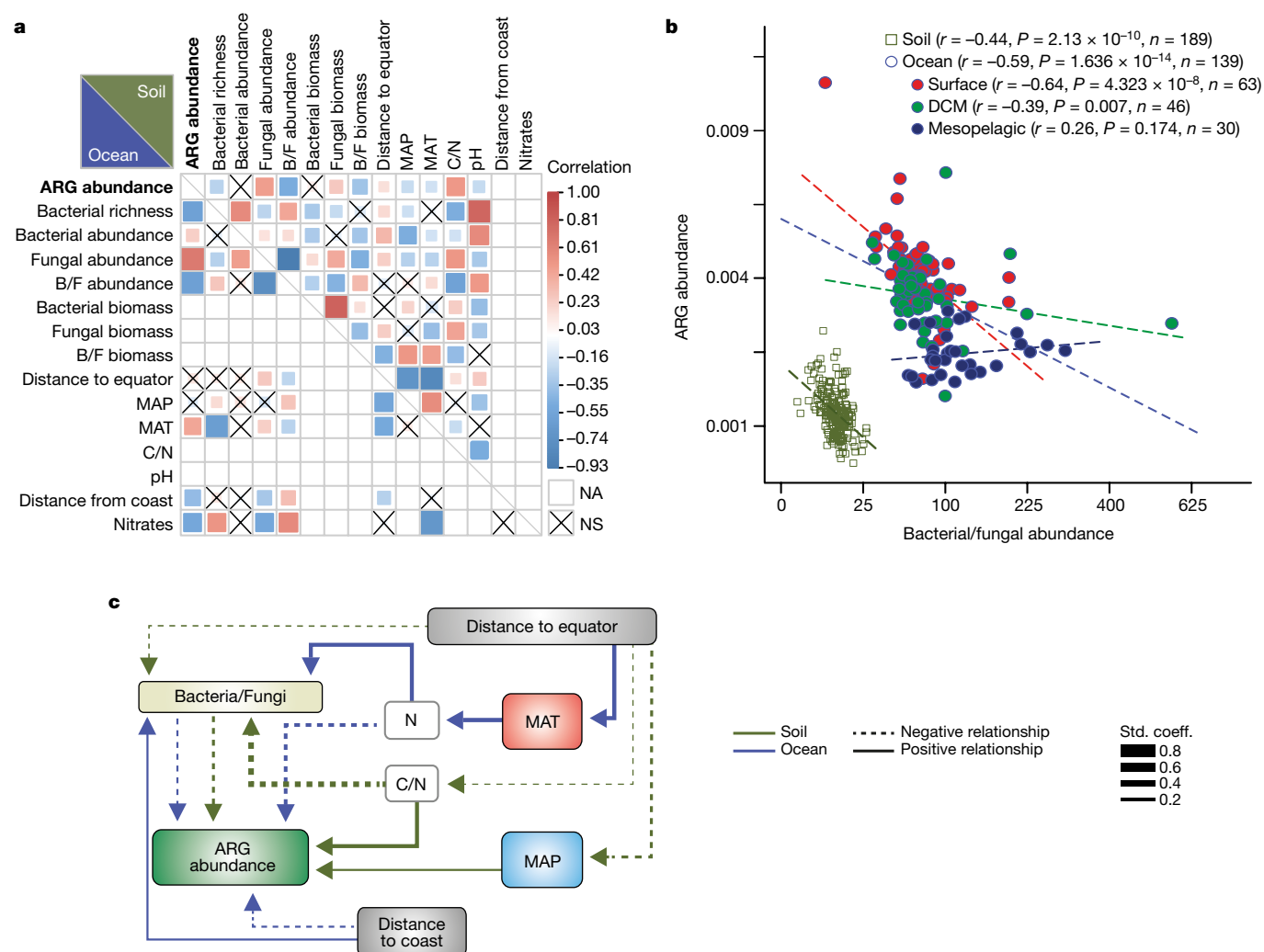


Fig. 2 | Global relative abundance of ARGs can be explained by a combination of biotic and abiotic factors. **a**, Pairwise Spearman's correlation matrix of the main biotic and abiotic determinants of the relative abundance of ARGs. **b**, Bacterial/fungal abundance ratio significantly correlated with the relative abundance of ARGs on a global scale. **c**, Structural equation modelling (SEM) of the relative abundance of ARGs in the soil (green) and ocean (blue) datasets (explaining 44% and 51% of variation, respectively; Supplementary Table 5). The goodness of fit was acceptable (soil dataset: root mean square error of estimation (RMSEA) = 0.00, P value for a test of close fit (P_{CLOSE}) = 0.989, n = 189

biologically independent samples; ocean dataset: RMSEA = 0.059, P_{CLOSE} = 0.302, n = 139 biologically independent samples). Abundance, relative abundance of miTags determined as fungi (including fungus-like protists) or bacteria; B/F, bacterial/fungal abundance or biomass ratio; bacterial richness, bacterial OTU (>97% similarity) richness on the basis of the metabarcoding dataset; biomass (nmol g⁻¹), absolute biomass on the basis of PLFA analysis; DCM, deep chlorophyll maximum; MAT, mean annual temperature; N, nitrates; NA, not applicable; NS, not significant ($P > 0.05$, $q > 0.1$); Std. coeff., standardized coefficients.

pattern of gene functional diversity across the latitudinal gradient, that is, niche differentiation, between bacteria and fungi (Fig. 1, Extended Data Fig. 2). Although increasing precipitation seems to favour higher fungal diversity, it is associated with higher bacterial/fungal biomass and abundance ratios (Extended Data Figs. 3d, g, 5f, h). The increasing proportion of fungi towards higher latitudes may be explained by competitive advantages, perhaps owing to a greater tolerance to nutrient and water limitation associated with potential long-distance transport by hyphae.

A role of inter-kingdom biotic interactions in determining the distributions of functional diversity and biomass in fungi and bacteria has been suggested previously²². As competition for resources affects the biomass of fungi and bacteria^{22,23}, we hypothesized that the bacterial/fungal biomass ratio is related to the prevalence of fungi and bacterial antibiotic-resistance capacity, because of broader activities of fungi than bacteria in using complex carbon substrates²⁴ as well as increased antibiotic production of fungi in high C/N environments²⁵. Consistent with this hypothesis, we found that both fungal biomass and the bacterial/fungal biomass ratio correlated with the relative abundance of ARGs (Extended Data Fig. 6) and that most fungal orthologous group

subcategories, particularly those involved in biosynthesis of antibiotic and reactive oxygen species, increased with soil C/N (Supplementary Table 4; Supplementary Discussion). We also found that the relative abundance of ARGs in topsoil is more strongly related to fungal relative abundance ($r = 0.435$, $P < 10^{-9}$) and bacterial/fungal abundance ratio ($r = -0.445$, $P < 10^{-12}$; Fig. 2b) than to bacterial relative abundance ($r = 0.232$, $P = 0.002$, on the basis of miTags), which is supported by our external validation dataset (fungal relative abundance $r = 0.637$, $P < 10^{-15}$; bacterial/fungal abundance ratio $r = -0.621$, $P < 10^{-15}$; bacterial relative abundance $r = 0.174$, $P = 0.036$). In addition, the relative abundance of ARGs in topsoil was significantly negatively correlated with bacterial phylogenetic diversity and OTU richness on the basis of the 16S rRNA gene (Spearman correlation, $P < 0.01$; Extended Data Figs. 7a, c, 8a), further supporting a role for biotic interactions in shaping microbial communities.

We also tested possible direct and indirect relationships between ARGs and 16 environmental predictors using structural equation modelling (SEM; Supplementary Table 5). The optimized model suggests that the soil C/N ratio and moisture, rather than pH—the predominant driver of bacterial diversity (Extended Data Fig. 3g, Supplementary

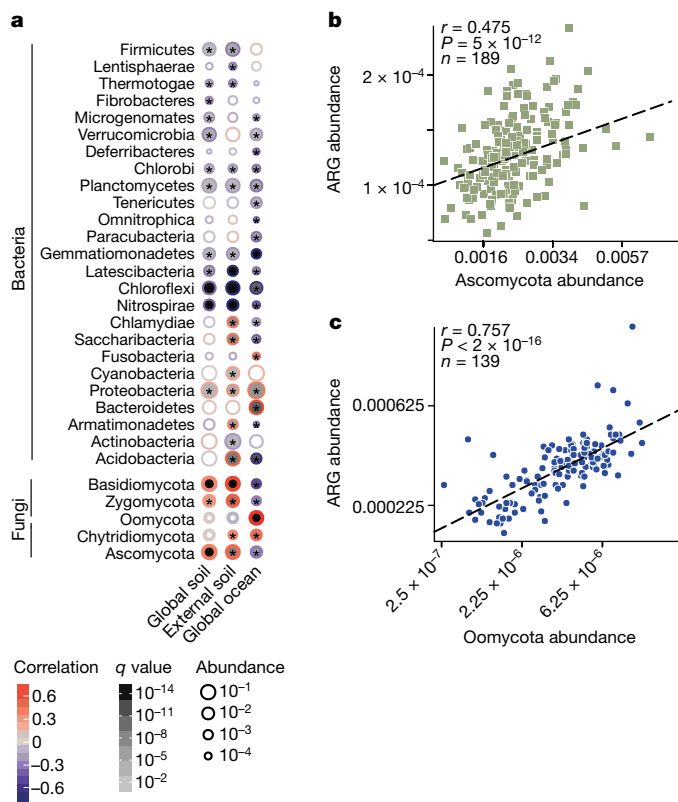


Fig. 3 | Fungi are the main determinants of the relative abundance of ARGs in soils and oceans. **a**, The association between the relative abundance of ARGs and major bacterial and fungal (including fungal-like protist) phyla in metagenomic samples from soils and oceans. Outer circle colour corresponds to the Pearson's correlation coefficient. Circle fill colour corresponds to the significance after adjustment for multiple testing (q value), as indicated in the legend. **b**, **c**, Relationships (non-parametric correlations) between the relative abundances of the most correlated fungal groups with ARGs in soil metagenomes from this study (**b**) and ocean metagenomes (**c**). For statistical details and significance, see Supplementary Table 8. Asterisks denote significance after Benjamini–Hochberg correction for multiple testing; $*q < 0.1$. See also Supplementary Discussion and Supplementary Table 8 for analogous results as in **a** but at the class level, and in other habitats besides soil and ocean including published non-forest and agricultural soil as well as human skin and gut samples.

Discussion)—affect the bacterial/fungal abundance ratio, which in turn affects the relative abundance of ARGs at the global scale (Fig. 2c). In line with increased production of antibiotics in high-competition environments, the soil C/N ratio was the best predictor for richness of fungal functional genes ($r^2 = 0.331$, $P < 10^{-15}$; Supplementary Table 3) and bacterial carbohydrate active enzyme (CAZyme) genes involved in degrading fungal carbohydrates ($r = 0.501$, $P < 10^{-12}$). The relative abundance of ARGs was also strongly correlated with C/N in the external validation dataset ($r = 0.505$, $P < 10^{-10}$).

Although the concomitant increase in antibiotic-resistance potential and the relative abundance of bacteria (as potential ARG carriers) was expected, the strong correlation of fungal relative abundance with the relative abundance of ARGs and in turn bacterial phylogenetic diversity may be explained by selection against bacteria that lack ARGs, such that bacteria surviving fungal antagonism are enriched for ARGs. Among all studied phyla, the relative abundance of Chloroflexi, Nitrospirae, and Gemmatimonadetes bacteria (on the basis of miTags), taxa with relatively low genomic ARG content (Supplementary Table 6) were most strongly negatively correlated with ARG relative abundance (Fig. 3a). By contrast, ARGs were strongly positively correlated with the relative abundance of Proteobacteria, which have the greatest average number of ARGs per genome²⁶ among bacteria (Supplementary Table 6), and the fungal phyla Ascomycota and Zygomycota *sensu lato*

(including Zoopagomycota and Mucoromycota) in both the global soil and the external validation datasets (Fig. 3a, b, Extended Data Fig. 9a, c, Supplementary Table 7). More specifically, ITS metabarcoding revealed increasing relative abundances of ARGs with numerous fungal OTUs (Supplementary Table 8), particularly those belonging to *Oidiodendron* (Myxotrichaceae, Ascomycota) and *Penicillium* (Aspergillaceae, Ascomycota), which are known antibiotic producers^{27,28} (Supplementary Discussion). Among bacterial ARGs, the relative abundance of efflux pumps and β -lactamases, which act specifically on fungal-derived antibiotics, were significantly correlated to the relative abundance of Ascomycota (Extended Data Fig. 10a, Supplementary Table 7). Actinobacteria, encompassing antibiotic-producing *Streptomyces*, also significantly correlated to ARG diversity in topsoil (Supplementary Table 6). Together these results suggest that relationships between organismal and ARG abundances are probably the result of selective and/or suppressive actions of antibiotics on bacteria.

Consistent with our observations in topsoil, we found evidence for antagonism between fungi and bacteria in oceans by reanalysing the distribution of ARGs in 139 water samples from the global *Tara* Oceans project¹³ (see Methods 'External metagenomic datasets'; Supplementary Table 1, Extended Data Fig. 8a): the fungus-like stramenopile class Oomycetes (water moulds) and the fungal phylum Chytridiomycota constituted the groups most strongly associated with the relative abundance of bacterial ARGs (Fig. 3a, c, Extended Data Figs. 9b, d, 10b, d). Although there is little direct evidence that oomycetes produce antibiotics, their high antagonistic activity can induce bacteria²⁹ and other organisms, including fungi³⁰, to produce antibiotics (Supplementary Discussion). As in topsoil, bacterial phylogenetic diversity was significantly negatively correlated with the relative abundance of ARGs in ocean samples (Extended Data Fig. 7b, c). In addition, the relative abundance of ARGs declined with increasing distance from the nearest coast in ocean samples (Extended Data Fig. 8b), which may reflect the effect of a decreasing nutrient gradient along distance from the coast on the pattern of bacterial and fungal abundance and in turn the abundance of ARGs. The agreement of results from these disparate habitats suggests that competition for resources related to nutrient availability and climate factors drive a eukaryotic–bacterial antagonism in both terrestrial and oceanic ecosystems.

Our results indicate that both environmental filtering and niche differentiation determine global soil microbial composition, with a minor role of dispersal limitation at this scale (for limitations, see Methods 'Metagenomics and metabarcoding analyses'). In particular, the global distributions of soil bacteria and fungi were most strongly associated with soil pH and precipitation, respectively. Our data further indicate that inter-kingdom antagonism, as reflected in the association of bacterial ARGs with fungal relative abundance, is also important in structuring microbial communities. Although further studies are needed to explicitly address the interplay between the bacterial/fungal abundance ratio and the abundance of ARGs, our data suggest that environmental variables that affect the bacterial/fungal abundance ratio may have consequences for microbial interactions and may favour fungi- or bacteria-driven soil nutrient cycling. This unprecedented view of the global patterns of microbial distributions indicates that global climate change may differentially affect bacterial and fungal community composition and their functional potential, because acidification, nitrogen pollution and shifts in precipitation all have contrasting effects on topsoil bacterial and fungal abundance, diversity and functioning.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0386-6>.

Received: 7 March 2017; Accepted: 13 June 2018;
Published online 1 August 2018.

1. Green, J. L., Bohannan, B. J. & Whitaker, R. J. Microbial biogeography: from taxonomy to traits. *Science* **320**, 1039–1043 (2008).
2. Reed, D. C., Algar, C. K., Huber, J. A. & Dick, G. J. Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proc. Natl Acad. Sci. USA* **111**, 1879–1884 (2014).
3. Maynard, D. S., Crowther, T. W. & Bradford, M. A. Fungal interactions reduce carbon use efficiency. *Ecol. Lett.* **20**, 1034–1042 (2017).
4. de Menezes, A. B., Richardson, A. E. & Thrall, P. H. Linking fungal–bacterial co-occurrences to soil ecosystem function. *Curr. Opin. Microbiol.* **37**, 135–141 (2017).
5. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
6. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
7. Rousk, J. et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).
8. de Boer, W., Folman, L. B., Summerbell, R. C. & Boddy, L. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol. Rev.* **29**, 795–811 (2005).
9. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51**, 933–938 (2001).
10. Green, J. & Bohannan, B. J. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507 (2006).
11. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE* **7**, e33865 (2012).
12. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
13. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
14. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
15. Logares, R. et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* **16**, 2659–2671 (2014).
16. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
17. Willig, M. R., Kaufman, D. & Stevens, R. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Syst.* **34**, 273–309 (2003).
18. Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
19. Baas-Becking, L. G. M. *Geobiologie; Of Inleiding tot de Milieukunde* (W. P. Van Stockum & Zoon NV, The Hague, 1934).
20. Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
21. Bryant, J. A., Stewart, F. J., Eppley, J. M. & DeLong, E. F. Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. *Ecology* **93**, 1659–1673 (2012).
22. Frey-Klett, P. et al. Bacterial–fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. *Microbiol. Mol. Biol. Rev.* **75**, 583–609 (2011).
23. Mille-Lindblom, C., Fischer, H. & Tranvik, J. L. Antagonism between bacteria and fungi: substrate competition and a possible tradeoff between fungal growth and tolerance towards bacteria. *Oikos* **113**, 233–242 (2006).
24. Koranda, M. et al. Fungal and bacterial utilization of organic substrates depends on substrate complexity and N availability. *FEMS Microbiol. Ecol.* **87**, 142–152 (2014).
25. Platas, G., Pelaez, F., Collado, J., Villuendas, G. & Diez, M. Screening of antimicrobial activities by aquatic hyphomycetes cultivated on various nutrient sources. *Cryptogam. Mycol.* **19**, 33–43 (1998).
26. Mende, D. R. et al. proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Res.* **45**, D529–D534 (2017).
27. Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J. Antibiot. (Tokyo)* **65**, 385–395 (2012).
28. Andersen, N. R. & Rasmussen, P. The constitution of clerocidin a new antibiotic isolated from *Oidiodendron truncatum*. *Tetrahedron Lett.* **25**, 465–468 (1984).
29. Zhao, Y., Qian, G., Chen, Y., Du, L. & Liu, F. Transcriptional and antagonistic responses of biocontrol strain *Lysobacter enzymogenes* OH11 to the plant pathogenic oomycete *Pythium aphanidermatum*. *Front. Microbiol.* **8**, 1025 (2017).
30. Takahashi, K. et al. Cladomarine, a new anti-saprolegniasis compound isolated from the deep-sea fungus, *Penicillium coralligerum* YK-247. *J. Antibiot. (Tokyo)* **70**, 911–914 (2017).

Acknowledgements The authors thank I. Liiv for technical and laboratory assistance; S. Waszak for comments on the manuscript; Y. P. Yuan and A. Glazek for bioinformatics support and A. Holm Viborg for help in retrieving the CAZY database. We also thank V. Benes, R. Hercog and other members of the EMBL GeneCore (Heidelberg), who provided assistance and facilities for sequencing. This study was funded by the Estonian Research Council (grants PUT171, PUT1317, PUT1399, IUT20-30, MOBERC, KIK, RMK, ECOLCHANGE), the Swedish Research Council (VR grant 2017-05019), Royal Swedish Academy of Sciences, Helge Axson Johnsons Stiftelse, EU COST Action FP1305 Biolink (STSM grant), Netherlands Organization for Scientific research (vidi grant 016.161.318), EMBL European Union's Horizon 2020 Research and Innovation Programme (#686070; DD-DeDaF) and Marie Skłodowska-Curie (600375).

Reviewer information *Nature* thanks S. Tringe and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions M.B., L.T. and P.B. conceived the project. L.T. supervised DNA extraction and sequencing. M.B., F.H., S.K.F., J.L.A., M.R. and P.M.B. designed and supervised the data analyses. F.H. designed and performed bioinformatics analysis. N.A.S. and P.A.O. performed biomass analysis. S.K.F., S.M., M.P., S.A., H.H., S.P., M.R.M., S.S. and L.T. contributed data. M.B., F.H., S.K.F., J.L.A., P.M.B., S.A., J.B.-P., M.H.M., L.P.C. and J.H.-C. performed the data analyses. M.B. wrote the first draft of the manuscript with input from F.H., S.F., J.L.A., L.T. and P.B. All authors contributed to data interpretation and editing of the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0386-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0386-6>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to M.B., L.T. or P.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Soil-sample preparation. Composite soil samples from 1,450 sites worldwide were collected using highly standardized protocols¹⁶. The sampling was conducted broadly across the most influential known environmental gradient (that is, the latitude) taking advantage of a global ‘natural laboratory’ to study the impact of climate on diversity across vegetation, biome and soil types and to enable testing of the effects of environmental parameters, spatial distance and biotic interactions in structuring microbial communities. We carefully selected representative sites for different vegetation types separated by spatial distances that were sufficient to minimize spatial autocorrelation and to cover most areas of the globe. Total DNA was extracted from 2 g of soil from each sample using the PowerMax Soil DNA Isolation kit (MoBio). A subset of 189 high-quality DNA samples representing different ecoregions spanning multiple forest, grassland and tundra biomes (Supplementary Table 1) were chosen for prokaryote and eukaryote metabarcoding (ribosomal rRNA genes) and whole metagenome analysis. Samples from desert ($n = 8$; G4010, G4034, S357, S359, S411, S414, S418 and S421) and mangrove ($n = 1$; G4023) biomes yielded sufficient DNA for metabarcoding, but not for metagenomics sequencing, thus these samples were used for global mapping of taxonomic diversity but excluded from all comparisons between functional and taxonomic diversity. One sample (S017) contained no 16S sequences; thus, altogether, 189 and 197 samples were used for metagenomics and metabarcoding analyses, respectively.

To determine the functional gene composition of each sample, 5 µg total soil DNA (300–400 bp fragments) was ligated to Illumina adaptors using the TruSeq Nano DNA HT Library Prep Kit (Illumina) and shotgun-sequenced in three runs of the Illumina HiSeq 2500 platform (2 × 250 bp paired-end chemistry, rapid run mode)³¹ in the Estonian Genomics Center (Tartu, Estonia). Taxonomic composition was estimated from the same DNA samples using ribosomal DNA metabarcoding for bacteria (16S V4 subregion) and eukaryotes (18S V9 subregion). For amplification of prokaryotes and eukaryotes, universal prokaryote primers 515F and 806RB³² (although this pair may discriminate against certain groups of Archaea and Bacteria such as Crenarchaeota/Thaumarchaeota (and SAR11³³) and eukaryote primers 1389f and 1510r³⁴ were used. Although the resolution of 16S rRNA sequencing is limited to assignments to the level of genus (and higher), it is currently a standard approach in profiling bacterial communities and thus enabled us at least to explore patterns at coarse phylogenetic resolution.

Each primer was tagged with a 10–12-base identifier barcode¹⁶. DNA samples were amplified using the following PCR conditions: 95 °C for 15 min, and then 30 cycles of 95 °C for 30 s, 50 °C for 45 s and 72 °C for 1 min with a final extension step at 72 °C for 10 min. The 25 µl PCR mix consisted of 16 µl sterilized H₂O, 5 µl 5 × HOT FIREPol Blend MasterMix (Solis Biodyne, Tartu, Estonia), 0.5 µl each primer (200 nM) and 3 µl template DNA. PCR products from three technical replicates were pooled and their relative quantity was evaluated after electrophoresis on an agarose gel. DNA samples producing no visible band or an overly strong band were amplified using 35 or 25 cycles, respectively. The amplicons were purified (FavourPrep Gel/PCR Purification Kit; Favourgen), checked for quality (ND 1000 spectrophotometer; NanoDrop Technologies), and quantified (Qubit dsDNA HS Assay Kit; Life Technologies). Quality and concentration of 16S amplicon pools were verified using Bioanalyzer HS DNA Analysis Kit (Agilent) and Qubit 2.0 Fluorometer with dsDNA HS Assay Kit (Thermo Fisher Scientific), respectively. Sequencing was performed on an Illumina MiSeq at the EMBL GeneCore facility (Heidelberg, Germany) using a v2 500 cycle kit, adjusting the read length to 300 and 200 bp for read1 and read2, respectively. 18S amplicon pools were quality checked using Bioanalyzer HS DNA Analysis Kit (Agilent), quantified using Qubit 2.0 Fluorometer with dsDNA HS Assay Kit (Thermo Fisher Scientific) and sequenced on an Illumina HiSeq at Estonian Genomics Center (Tartu, Estonia). Sequences resulting from potential contamination and tag switching were identified and discarded on the basis of two negative and positive control samples per sequencing run.

Soil chemical analysis and biomass analysis. All topsoil samples were subjected to chemical analysis of pH_{KCl}, P_{total} (total phosphorus), K, Ca and Mg; the content of ¹²C, ¹³C, ¹⁴N and ¹⁵N was determined using an elemental analyser (Eurovector) coupled with an isotope-ratio mass spectrometer³⁵.

To calculate the absolute abundance of bacteria and fungi using an independent approach, bacterial and fungal biomass were estimated from PLFAs³⁶ in nmol g⁻¹ as follows. Lipids were extracted from 2 g freeze dried soil in a one-phase solution of chloroform, methanol and citrate buffer³⁷. Chloroform and citrate buffer was added to split the collected extract into one lipophilic phase, and one hydrophilic phase. The lipid phase was collected and applied on a pre-packed silica column³⁷. The lipids were separated into neutral lipids, intermediate lipids and polar lipids (containing the phospholipids) by subsequent elution with chloroform, acetone

and methanol. The neutral and phospholipids were dried using a speed vac. Methyl nonadecanoic acid (Me19:0) was added as an internal standard. The lipids were subjected to a mild alkaline methanolysis, in which fatty acids were derivatized to fatty acid methyl esters (FAMES). The FAMES from neutral (NLFAs) and phospholipids (PLFAs) were dried, using speed vac, and then dissolved in hexane before analysis on a gas chromatograph as described³⁸. Fungal biomass was estimated as the concentration of PLFA 18:2ω6,9 and bacterial biomass from the sum of nine PLFAs (i15:0, i16:0, i17:0, a15:0, a17:0, cy17:0, cy19:0, 10Me17:0 and 10Me18:0)³⁷. The nomenclature of fatty acids was according to previously published work³⁸.

Acquisition of metadata from public databases. Climate data including monthly temperature and precipitation were obtained from the WorldClim database (<http://www.worldclim.org>). In addition, estimates of soil carbon, moisture, pH, potential evapotranspiration (PET) and net primary productivity (NPP) at 30 arc minute resolution were obtained from the Atlas of the Biosphere (<https://nelson.wisc.edu/sage/data-and-models/atlas/maps.php>). Samples were categorized into 11 biomes⁹, with all grassland biomes being categorized as ‘grasslands’. Thus, the following biomes were considered and summarized to three global levels: moist tropical forests, tropical montane forests and dry tropical forests, savannahs as tropical; Mediterranean, grasslands and shrublands, southern temperate forests, coniferous temperate forests and deciduous temperate forests as temperate; and boreal forests and arctic tundra as boreal–arctic. The time from the last fire disturbance was estimated on the basis of enquiries to local authorities or collaborators and evidence from the field.

Metagenomics and metabarcoding analyses. *Processing of metagenomics sequence data.* Most soil microorganisms are uncultured, making their identification difficult. Metagenomics analysis has emerged as a way around this to capture both genetic and phylogenetic diversity. As such, it can only directly reveal the potential for functions through determining and tracing gene family abundances (as opposed to realized protein activity), which may be involved in various functional pathways³⁹, but we can safely assume a strong correspondence between gene functional potential and the resulting ecosystem functioning⁴⁰ or enzyme activities⁴¹.

Reads obtained from the shotgun metagenome sequencing of topsoil samples were quality-filtered, if the estimated accumulated error exceeded 2.5 with a probability of ≥ 0.01 ⁴², or >1 ambiguous position. Reads were trimmed if base quality dropped below 20 in a window of 15 bases at the 3' end, or if the accumulated error exceeded 2 using the sdm read filtering software⁴³. After this, all reads shorter than 70% of the maximum expected read length (250 bp unless noted otherwise for external datasets) were removed. This resulted in retention of 894,017,558 out of 1,307,037,136 reads in total (Supplementary Table 1). We implemented a direct mapping approach to estimate the functional gene composition of each sample. First, the quality-filtered read pairs were merged using FLASH⁴⁴. The merged and unmerged reads were then mapped against functional reference sequence databases (see below) using DIAMOND v0.8.10 in blastx mode using ‘-k 5 -e 1e-4 --sensitive’ options. The mapping scores of two unmerged query reads that mapped to the same target were combined to avoid double counting. In this case, the hit scores were combined by selecting the lower of the two *e* values and the sum of the bit scores from the two hits. The best hit for a given query was based on the highest bit score, longest alignment length and highest percentage identity to the subject sequence. Finally, aligned reads were filtered to those that had an alignment percentage identity >50% and $e < 1 \times 10^{-9}$ (see ‘Parameterization and validation of metagenomics approach’ for parameter choice).

The functional databases to which metagenomic reads were mapped included gene categories related to ROS sources (peroxidases genes databases^{45,46}, KEGG⁴⁷ (Kyoto Encyclopedia of Genes and Genomes) and CAZyme genes (<http://www.cazy.org>, accessed 22 November 2015)⁴⁸. To facilitate the interpretation of the results, the relative abundance of CAZyme genes were summed on the basis of the substrates for each gene family. Substrate utilization information for CAZyme families was obtained from previously published work^{49,50} as well as the CAZypedia (http://www.cazypedia.org/index.php?title=Carbohydrate-binding_modules&oldid=9411). On the basis of the KEGG orthologue abundance matrices we calculated SEED functional module abundances. For functional annotations of metagenomic reads, we used in silico annotation on the basis of a curated database of the orthologous gene family resource eggNOG 4.5¹⁴.

For all databases that included taxonomic information (eggNOG, KEGG, CAZY), reads were mapped competitively against all kingdoms and assigned into prokaryotic and eukaryotic groups, on the basis of best bit score in the alignment and the taxonomic annotation provided within the database at kingdom level. All functional abundance matrices were normalized to the total number of reads used for mapping in the statistical analysis, unless mentioned otherwise (for example, rarefied in the case of diversity analysis, see ‘Statistical analyses’). This normalization better takes into account differences in library size as it has the advantage of including the fraction of unmapped (that is functionally unclassified) reads. Although there are limitations to using relative abundance of genes, our analysis shows which potential functions are relatively more important. Without any

normalization, such analyses cannot be performed. It is currently difficult to test the absolute numbers, owing to limitations of reliably quantifying soil DNA resulting from differences in extraction efficiency and the level of degradation.

To identify ARGs in our metagenome samples, the merged and unmerged reads were mapped to a homology expansion⁵¹ of the Antibiotic Resistance Gene Data Base (ARDB). Only hits that passed the minimum sequence identity values as listed in the ARDB for each family were taken further into account. Although newer ARG databases exist, only the ARDB presently has curated family inclusion thresholds that directly allow application to our topsoil dataset: as soil microbial diversity is so large, unlike for gut datasets, high-fidelity gene catalogue construction will not be possible until many more samples are available. Therefore, direct mapping of reads to the gene family databases becomes necessary for our analysis, in turn necessitating ARG inclusion thresholds that are well-defined for single reads, not merely for full-length genes. Thus, the cut-offs curated by ResFams⁵² or CARD⁵³, for example, are inappropriate, as they are defined in the length-dependent bit-score space. The ARDB cut-offs, however, are defined as sequence identities and thus in principle are applicable to sequences shorter than full length. Because of these technical limitations, we used a soil-gene catalogue to determine CARD-based ARG abundance matrices (see 'Gene catalogue construction').

It is important to note that measurements of functional genes, including ARGs, represent relative proportions of different gene families, because the absolute amount of DNA differs among samples. This necessitates the use of statistical tests that do not assume absolute measurements, and centres analysis of this type on comparisons across the set of samples.

Estimation of taxa abundance using miTag. We used a miTag approach¹⁵ to determine bacterial and fungal community composition from metagenome sequence data. First, SortMeRNA⁵⁴ was used to extract and blast search rRNA genes against the SILVA LSU/SSU database. Reads approximately matching these databases with $e < 10^{-1}$ were further filtered with custom Perl and C++ scripts, using FLASH to attempt to merge all matched read pairs. In case read pairs could not be merged, which happens when the overlap between read pairs is too small, the reads were interleaved such that the second read pair was reverse complemented and then sequentially added to the first read. To fine-match candidate interleaved or merged reads to the SILVA LSU/SSU databases, lambda⁵⁵ was used. Using the lowest common ancestor (LCA) algorithm adapted from LotuS (v.1.462)⁴³, we determined the identity of filtered reads on the basis of lambda hits. This included a filtering step, in which queries were only assigned to phyla and classes if they had at least 88% and 91% similarity to the best database hit, respectively. The taxon-by-sample matrices were normalized to the total number of reads per sample to minimize the effects of uneven sequencing depth. The average of SSU and LSU matrices was used for calculating the relative abundance of phyla or classes. The abundance of miTag sequences matching bacteria and fungi was used to determine the bacterial/fungal abundance ratio. Although LSU/SSU assessments refer to the number of fungal cells rather than the number of discrete multicellular fungi (as this can apply to all samples equally), it is not systematically biased for comparing the trends of bacterial to fungal abundance across samples.

External metagenomic datasets. To validate and compare the global trends at smaller scales, we used a regional scale dataset of 145 topsoil samples that was generated and processed using the same protocol as our global dataset (Supplementary Table 1).

In addition, to compare patterns of ARG diversity in soils and oceans on a global scale, we re-analysed the metagenomics datasets of the *Tara* Oceans¹³, including all size fractions (Supplementary Table 1). After quality filtering, 41,790,928,650 out of 43,076,016,494 reads were retained from the *Tara* Oceans dataset.

The quality-filtered reads from all datasets were mapped to the corresponding databases using DIAMOND, with the exception that no merging of read pairs was attempted, because the chances of finding overlapping reads were too low (with a read length of 100 bp and insert size of 300 bp (*Tara* Oceans)). Sequences for SSU/LSU miTags were extracted from these metagenomics datasets as described above. ARG abundance matrices were also obtained from the *Tara* Oceans project on the basis of the published gene catalogues annotated using a similar approach as in the current study.

Gene catalogue construction. To create a gene catalogue, we first searched for complete reference genes that matched to read pairs in our collection using Bowtie2⁵⁶ with the options '--no-unal --end-to-end'. The resulting bam files were sorted and indexed using samtools 1.3.1⁵⁷ and the jgi_summarize_bam_contig_depths provided with MetaBat⁵⁸ was used to create a depth profile of genes from the reference databases that were covered with $\geq 95\%$ nucleotide identity. This cut-off is commonly used in constructing gene catalogues^{13,59} and chosen to delineate genes belonging to the same species. Using the coverage information, we extracted all genes that had at least 200 bp with $\geq 1\times$ coverage by reads from our topsoil metagenomes. The reference databases included an ocean microbial gene catalogue¹³, a gut microbial gene catalogue¹², as well as all genes extracted from 25,038 published bacterial genomes²⁶. Altogether, 273,723, 2,376 and 8,642 genes

from proGenomes, IGC and Tara database, respectively, could be matched to soil reads and were used in the gene catalogue.

The majority of genes in our catalogue were assembled from the topsoil samples presented here. To reduce the likelihood of chimaeric reads, each sample was assembled separately using Spades 3.7-0 (development version obtained from the authors)⁶⁰ in metagenomic mode with the parameters '--only-assembler -m 500 --meta -k 21,33,67,111,127'. Only sdm-filtered⁴³ paired reads were used in the assembly, with the same read-filtering parameters as described above. Resulting assemblies had an average N50 of 469 bases (total of all assemblies 21,538 Mb). The low N50 reflects difficulties in the assembly of soil metagenomes, which probably reflects the vast microbial genetic diversity of these ecosystems. We further de novo assembled reads from two other deep sequencing soil⁶¹ and sediment studies⁶², using the same procedure and parameters, except that the Spades parameter '-k 21,33,67,77' was adjusted to a shorter read length. Furthermore, we included publicly available data from the European Nucleotide Archive (ENA). The ENA was queried to identify all projects with publicly available metagenomes and whose metadata contained the keyword 'soil'. The initial set of hits was then manually curated to select relevant project and/or samples that were assembled as described above. Additionally, we integrated gene predictions from soil metagenomes downloaded from MG-RAST⁶³ (Supplementary Table 1). Assembly was not attempted for these samples owing to the absence of paired-end reads, and relatively low read depth; rather, only long reads or assemblies directly uploaded to MG-RAST with ≥ 400 bp length were retained. Therefore, only scaffolds and long reads, with at least 400 bp length, were used for analysis. On these filtered sequences, genes were de novo predicted using prodigal 2.6.1⁶⁴ in metagenomic mode. Finally, we merged the predicted genes from assemblies, long reads, gene catalogues and references genomes to construct a comprehensive soil gene catalogue.

Thus, 53,294,555,100 reads were processed, of which 31,015,827,636 (58.20%) passed our stringent quality control. The initial gene set predicted on the soil assemblies and long reads was separated into 17,114,295 complete genes and 111,875,596 incomplete genes. A non-redundant gene catalogue was built by comparing all genes to each other. This operation was performed initially in amino-acid space using DIAMOND⁶⁵. Subsequently, any reported hits were checked in nucleotide space. Any gene that covered at least 90% of another one (with at least 95% identity over the covered area) was considered to be a potential representative of it (genes are also potential representatives of themselves). The final set was chosen by greedily picking the genes that were representative of the highest number of input genes until all genes in the original input have at least one representative in the output. This resulted in a gene catalogue with a total of 159,907,547 non-redundant genes at 95% nucleotide identity cut-off. We mapped reads from our experiment onto the gene catalogue with bwa⁶⁶, requiring >45 nt overlap and $>95\%$ identity. The average mapping rate was $26.2 \pm 7.4\%$. Although the gene catalogue is an invaluable resource for future explorations of the soil microbiome, we decided to rely on using the direct mapping approach to gene functional composition, owing to the low overall mapping rate. Furthermore, using minimap2⁶⁷ to find genes at 95% similarity threshold, we compared the soil gene catalogue with the *Tara* Oceans gene catalogue¹³, human gut gene catalogue¹² and the proGenomes prokaryotic database²⁶. The gene catalogue nucleotide and amino acid sequences and abundance matrix estimates from rtk⁶⁸ have been deposited at http://vm-lux.embl.de/~hildebra/Soil_gene_cat/.

Estimation of ARG abundance using CARD. CARD abundances in topsoil samples were estimated by annotating the soil gene catalogue using a DIAMOND search of the predicted amino acid sequences against the CARD database and filtering hits to the specified bit-score cut-offs in the CARD database. On the basis of gene abundances in each sample, we estimated the abundance of different CARD categories per metagenomic sample. Despite qualitative similarities in overall trends of ARDB and CARD abundance matrices, CARD abundance estimation is limited by being based on the gene catalogue (only a $26.2 \pm 7.4\%$ of all metagenomic reads could be mapped to the gene catalogue).

Processing of metabarcoding sequence data. The LotuS pipeline was used for bacterial 16S rRNA amplicon sequence processing. Reads were demultiplexed with modified quality-filtering settings for MiSeq reads, increasing strictness to avoid false positive OTUs. These modified options were the requirement of correctly detected forward 16S primer, trimming of reads after an accumulated error of 1 and rejecting reads below 28 average quality or, exceeding an estimated accumulated error >2.5 with a probability of ≥ 0.01 ⁴². Furthermore, we required each unique read (reads preclustered at 100% identity) to be present eight or more times in at least one sample, four or more times in at least two samples, or three or more times in at least three samples. In total 27,883,607 read pairs were quality-filtered and clustered with uparse⁶⁹ at 97% identity. Chimeric OTUs were detected and removed on the basis of both reference-based and de novo chimaera checking algorithms, using the RDP reference database (http://drive5.com/uchime/rdp_gold.fa) in uchime⁶⁹, resulting in 13,070,436 high-quality read pairs to generate and estimate the abundance of bacterial OTUs. The seed sequence for each OTU cluster was

selected from all read pairs assigned to that OTU, selecting the read pair with the highest overall quality and closest to the OTU centroid. Selected OTU seed read pairs were merged with FLASH⁴⁴ and a taxonomic identity was assigned to each OTU by aligning full-length sequences with lambda to the SILVA v.123 database⁷⁰ and the LotuS least common ancestor (LCA) algorithm. This was performed using the following LotuS command line options: '-p miSeq -derepMin 8:1,4,2,3:3 --simBasedTaxo 2 --refDB SLV -thr 8'. OTU abundances per sample were summed to class and phylum level per sample, according to their taxonomic classification, to obtain taxa abundance matrices. However, the choice of clustering method (for example, Swarm) and identity threshold had little effect on retrieved OTU richness (comparison with 99% threshold: $r = 0.977$, $P < 10^{-15}$; comparison with Swarm clustering: $r = 0.979$, $P < 10^{-15}$).

For eukaryotic 18S rRNA genes, we used the same options in LotuS, except that reads were rejected if they did not occur at least six times each in a minimum of two samples or at least four times each in a minimum of three samples. This was done to account for lower sequencing depth in 18S rRNA compared to 16S rRNA dataset. Furthermore, the database to annotate fungal taxonomy was extended to include general annotations of SILVA and information from unicellular eukaryotes (PR² database⁷¹). Of 7,462,813 reads, 2,890,093 passed quality filtering. The fungal ITS metabarcoding dataset¹⁶ was downloaded and used in addition to 18S data in specific analyses, such as finding fungal OTUs associated with ARG relative abundance. The resulting taxon abundance matrix was further filtered to remove sequences of chloroplast origin for all three metabarcoding experiments.

Full-length sequences representing OTUs were aligned using the SILVA reference alignment as a template in mothur⁷². A phylogenetic tree was constructed using FastTree⁷³ with the maximum-likelihood method using default settings. This program uses the Jukes–Cantor models to correct for multiple substitutions. *Parameterization and validation of metagenomics approach.* Although we used state-of-art molecular approaches, there are several potential limitations regarding our analyses related to the technologies used. All metagenomics and amplicon-based analyses are affected by taxonomic biases in sequence databases, whereas (PCR-free) miTag as well as amplicon sequencing are biased owing to differential ribosomal gene copy number across taxonomic groups. Amplicon-based metabarcoding, specifically, is affected by both primer PCR artefacts and PCR biases that may affect estimates of absolute organism abundance. These biases are inherent to all metagenomics and metabarcoding studies. However, all these biases affect different samples equally (same rRNA gene copy numbers, same PCR biases per species, same database bias per taxa) and thus we estimate that our results are robust to these methodological shortcomings. Shotgun-based metagenomics is affected by reference bias, in which human pathogens or Proteobacteria are overrepresented. The necessity for lenient thresholds becomes obvious from annotating phylogenetic profiles with MetaPhlAn2⁷⁴ using standard parameters: whereas we observed that most fungal phyla are present abundantly in our samples, MetaPhlAn2 detected Ascomycota in only 2 out of 189 samples. In 48 out of 189 samples, no organism (bacteria/archaea/eukaryotes) was detected, and the most abundant phylum was Proteobacteria (55%). As these results are clearly deviating from our miTag, 16S, 18S and ITS analyses, specific database cut-off thresholds were required for this project.

To optimize the analysis pipeline and identify suitable e values for filtering blastx results, we used metagenomic simulations of four reference genomes for which CAZy assignments in the CAZy database were available. Simulated reads were created as 250-bp paired reads with 400 bp insert at differing sequence abundances from the four reference genomes in each simulated metagenome, using iMessi⁷⁵. For this simulated dataset, we used the pipeline described above to derive CAZy functional profiles. We found that querying short reads processed as above against databases results in the retrieval of most genes at relative abundances consistent with expectations on the basis of the reference genomes at $e < 10^{-9}$ ($r = 0.95 \pm 0.01$, $P < 0.001$). Furthermore, we simulated 200 metagenomes from 18 bacterial genomes, five bacterial plasmids, one fungal mitochondrion and two fungal genomes at differing relative proportions in each of these simulated metagenomes (Supplementary Table 11). We subsequently simulated 1,000,000 reads of 250-bp and 400-bp insert size using iMessi, and mapped these against reference databases and retained hits that fulfilled the following arbitrary criteria (used in all subsequent analyses): e value cut-off of 10^{-9} , alignment length ≥ 20 amino acids, and similarity $\geq 50\%$ amino acids to the target sequence. From these, we generated functional profiles and found a strong correlation of simulated to expected functional metagenomic composition on the basis of mixed fungal and bacterial genomes ($r = 0.94 \pm 0.05$, $P < 0.001$).

Estimating fungal antibiotics production. We also specifically screened for fungal gene clusters directly associated with antibiotic activity, on the basis of a compiled database of MiBiG (minimum information about a biosynthetic gene cluster, <https://mibig.secondarymetabolites.org>) repository entries that describe gene clusters for which the products have been shown experimentally to display antimicrobial activities (Supplementary Table 12). To extend the range of genes

that can be associated with the validated, antibiotic-producing, MiBiG protein domains, we downloaded all published non-redundant fungal genomes deposited in JGI (Supplementary Table 14) as well as all non-redundant fungal genes deposited in NCBI. The set of MiBiG and fungus-derived genes was screened with custom hidden Markov models for domains from secondary metabolite production (specifically these were dmat, AMP-binding, Condensation, PKS_KS and Terpene synthesis domains). All identified domains were aligned together with the MiBiG domains using Clustal Omega⁷⁶ and a tree was constructed with FastTree2. Phylogenetic trees were rooted to midpoint and automatically scanned to identify highly supported clades (aLRT branch support ≥ 0.99) in which antibiotic-producing MiBiG domains were monophyletically grouped. The average nucleotide identity within each such group was subsequently used as identity cut-off in the mapping step. All metagenomic reads were mapped with DIAMOND in blastx mode to the newly created database, using the previously mentioned sequence identity cut-offs and rejecting domains of reads that were mapping to bacterial non-supervised orthologous groups.

Statistical analyses. *Data normalization and diversity estimates.* All statistical analyses were performed using specific packages in R (v.3.3.2) unless otherwise noted. Diversity parameters were estimated from OTU and functional gene matrices that were rarefied to an equal number per sample to reduce the effect of variation in sequencing depth using the function `rarefy` in `vegan` (v.2.2.1; <https://cran.r-project.org/web/packages/vegan/index.html>). ARG matrices were normalized to the total number of merged and singleton reads. The total abundance of ARGs per sample was estimated by summing the abundance of all individual ARGs per sample. ARG diversity measures indicate the variety and their proportions produced.

From the rarefied matrices we calculated OTU, orthologous group and CAZyme gene richness (function `specnumber`) and diversity (function `diversity` on the basis of the inverse Simpson index). The latter measure accounts for both richness and evenness, and it gives more weight to abundant groups compared to the Shannon index. Our results were robust to choice of index, and the various diversity indices were highly correlated in the present dataset (for example, bacterial taxonomic diversities calculated using inverse Simpson versus using Shannon diversity were highly correlated: $r = 0.888$, $P < 10^{-15}$; for a comparison of richness and diversity trends, see Extended Data Fig. 2b, c). As evenness and richness were highly correlated in all datasets, we report the results that, on the basis of the diversity index, represent both richness and evenness. The rarefaction process was repeated for calculating taxonomic and gene functional diversity and richness on the basis of the average of 100 rarefied datasets.

Phylogenetic diversity was calculated on the basis of Faith's Phylogenetic Diversity (PD) metric⁷⁷ in the Picante package (v.1.6-2; <https://cran.r-project.org/web/packages/picante/index.html>). In addition, to assess phylogenetic clustering and overdispersion, nearest relative index (NRI) and nearest taxon index (NTI) were calculated in Picante. Although both measures are closely related, NRI is more sensitive to phylogenetic diversity at deep nodes, whereas NTI is more sensitive to phylogenetic clustering towards tips. A null model of shuffling taxon labels (100 times) was used to randomize phylogenetic relationships among OTUs.

Correlating environmental parameters to taxa and functions. To identify the main determinants of taxonomic and gene functional composition or diversity and relative abundance of phyla and classes, we used a series of statistical tests. We included all prominent environmental variables that we expected to have a significant effect on microbial diversity on the basis of previous studies, and which were feasible to collect. These included soil pH, carbon and nutrient levels and factors that can affect these, such as fire, assuming soil as the major resource for microbial nutrition. We also included isotope ratios of nitrogen ($\delta^{15}\text{N}$) and carbon ($\delta^{13}\text{C}$) as these provide principal components for carbon and nitrogen cycling. To avoid overfitting and to ensure model simplicity, we excluded the variables that had no significant effect on fungal or bacterial diversity, such as altitude, age of vegetation, plant diversity and community (the first two principal component analysis axes of plant community variation at both genus and family level) and basal areas of trees. Thus, for univariate regression modelling, 16 variables (Supplementary Table 14) were included.

To understand which factors explain the orthologous group- and OTU-based community composition, variable selection was performed in the `Forward.sel` function of `Packfor` (v.0.0-8/r109; https://r-forge.r-project.org/R/?group_id=195) according to the coefficient of determination (threshold, $r^2 = 0.01$). All functional and taxonomic compositional matrices were transformed using Hellinger transformation before statistical analysis. Furthermore, Mantel tests and partial Mantel tests were used to test the effects of geographical versus environmental distances on the compositional similarity of OTUs and orthologous groups as implemented in `vegan`. Mantel tests allow testing of the correlation between two distance matrices, partial Mantel tests are similar but also control for variation in a third distance matrix. In our analysis, we controlled for the effect of geographic distance while testing the correlation of environmental variation and functional or taxonomic composition variation. The importance of biome type in explaining functional gene and taxonomic composition was tested in permutational multivariate

analysis of variance (PERMANOVA) using the Adonis function of *vegan* (using 10³ permutation for calculating pseudo-*F* test statistic and its statistical significance). For constructing orthologous group and OTU distance matrices, the Bray–Curtis dissimilarity was calculated between each pair of samples. Great-circle distance was used to calculate a geographic distance matrix between samples on the basis of geographical coordinates. This test compares the intragroup distances to intergroup distances in a permutation scheme and from this assesses significance. PERMANOVA post hoc *P* values were corrected for multiple testing using the Benjamini–Hochberg correction. We visualized taxonomic (OTU) and functional (orthologous group) composition of bacteria using global nonmetric multidimensional scaling (GNMDS) in *vegan* with the following options: two dimensions, initial configurations = 100, maximum iterations = 200 and minimum stress improvement in each iteration = 10^{−7}. The main environmental drivers of the relative abundance of major taxonomic groups and main functional categories were recovered by random forest analysis⁷⁸ using the R package *randomForest* (v.4.6-10; <https://cran.r-project.org/web/packages/randomForest/index.html>).

To examine latitudinal gradients of diversity at phylum level (Fig. 2), the diversity of OTUs assigned to each phylum was calculated on the basis of inverse Simpson index. Diversity values were modelled in response to environmental variables and predicted values were extracted, which were used in a clustering and bootstrapping analysis to depict the similarities of phyla environmental associations using *pvclust* (v.1.3-2; <https://cran.r-project.org/web/packages/pvclust/index.html>) with 1,000 iterations. To model latitudinal gradients and environmental associations of diversity and biomass (Fig. 1, Extended Data Fig. 3), we compared the goodness of fit estimates between first and second order polynomial models on the basis of the corrected Akaike information criterion (AICc) using analysis of variance (ANOVA). AICc reflects both goodness of fit and parsimony of the models.

For univariate regression modelling of diversity and biomass measures, ordinary least squares (OLS) or generalized least squares (GLS) regression models were used depending on the importance of the spatial component in the nlme package (v.3.1-120; <https://cran.r-project.org/web/packages/nlme/index.html>). The model variance structure (Gaussian, exponential, spherical and linear) was evaluated on the basis of AICc. After selection of variance structure, variables were combined in a set of models with specified variance structure (that is, the number of tested models = 2^{number of variables}). The resulting models were sorted according to AICc values to reveal the best model. Lists of the five best-fitting models for each response variable are given in Supplementary Table 4. Prior to model selection, all variables were evaluated for linearity, normality, and multicollinearity (excluded if the variance inflation factor was >5). The degree of polynomial functions (linear, quadratic, cubic) was chosen on the basis of the lowest AIC values. Because of nonlinear relationships with response variables, a quadratic term for pH was also included in the model selection procedure. The accuracy of the final models was evaluated using tenfold ‘leave-one-out’ cross-validation. For this, we used 1,000 randomly sampled 90%-data subsets for model training and predicting the withheld data. To minimize biases owing to the partitioning of the data and potential overfitting, the average of 1,000 resulting determination coefficients are reported as cross-validated *r*² (*r*² cv.) for each regression model.

Correlating biotic interactions to taxa and functions. To test the associations of biotic variables on ARG relative abundance, we used a sparse partial least squares (sPLS) analysis, which reduces dimensionality by projecting predictor variables onto latent components to identify the 16S/18S lineages (phyla and classes) and the ITS OTUs most strongly associated with ARG relative abundance, as implemented in the *mixOmics* (v.5.0-4; <https://cran.r-project.org/web/packages/mixOmics/index.html>) package. ARG composition and taxonomic community matrices (miTags classes and phyla and ITS OTUs) were normalized to library size using Hellinger transformation. Significance of associations was examined by bootstrap tests of subsets of each dataset. We subsequently used partial least squares (PLS) analysis to predict ARG relative abundance on the basis of significantly correlated lineages, which allows the dimensionality of multivariate data to be reduced into PLS components. Optimal numbers of PLS components for prediction of the relative ARG abundance were selected on the basis of leave-one-out cross-validation. To confirm the results of PLS analysis, we further used a cross-validated LASSO model to simultaneously perform variable selection and model fitting, as implemented in *glmnet* (v.2.0-2; <https://cran.r-project.org/web/packages/glmnet/index.html>). First the lambda shrinkage parameter was determined from a cross-validated LASSO-penalized logistic regression classifier. Using this shrinkage parameter, a new logistic regression classifier was fit to the data to predict ARG relative abundance.

To further test direct and indirect effects of geographic and environmental variables on microbial distributions, we built SEM models in the AMOS software (SPSS) by including predictors of the best GLS model. In a priori models, all indirect and direct links between variables were established on the basis of their pairwise correlations. We subsequently removed non-significant links and variables or created new links between error terms until a significant model fit was achieved. Goodness of fit was assessed on the basis of a χ^2 test to evaluate

the difference between the observed and estimated by-model covariance matrices (a non-significant value indicates that the model fits the observed data). We also used RMSEA and *P*_{CLOSE} to assess the discrepancy between the observed data and model per degree of freedom, which is less sensitive to sample size compared to the χ^2 test (RMSEA < 0.08 and *P*_{CLOSE} > 0.05 show a good fit). Observed correlations between diversity and environmental values can serve as the first step towards understanding the structure and function of global topsoil microbiome, however, they are not proof of causations and mechanism. Despite the fact that we used SEM modelling to infer indirect links, we cannot preclude the possibility of other biotic or soil variables confounded with climate variables that we did not include in our models. Further laboratory experiments may be able to address the causality of relationships reported in this study.

Differences between univariate variables such as taxonomic and functional richness were tested using a non-parametric Wilcoxon rank-sum test, with Benjamini–Hochberg multiple testing correction. Post hoc statistical testing for significant differences between all combinations of two groups was conducted only for taxa with *P* < 0.2 in the Kruskal–Wallis test. For this, Wilcoxon rank-sum tests were calculated for all possible group combinations and corrected for multiple testing using Benjamini–Hochberg multiple testing correction.

Geographic coordinates were plotted on a world map transformed to a Winkler2 projection, using the *maptools* package (v.0.8-36; <https://cran.r-project.org/web/packages/maptools/index.html>).

Limitations of statistical modelling on a global scale. Although we performed cross-validations to test the accuracy of most of our statistical models, predictions might be limited by the vast diversity in soil microbiomes. For example, strong local variation in soil pH may lead to deviation from general patterns, which is a common limitation in environmental sciences. Given the large spatial scale and strong environmental gradient in our sampling design, and long-term persistence of DNA in soil⁷⁹, seasonal variation in soils is expected to have a minor impact⁸⁰ (in contrast to the oceans). In addition, the vast majority of our samples were collected during the growing season, further reducing possible seasonal biases. We nevertheless tested the effect of sampling month and seasons and found no significant effect of seasonality on diversity indices (*P* > 0.05). We also compared the effect of seasons and years in a time series study in two of our sites, which revealed no seasonal effects on richness and composition (unpublished data). In particular, the relationship between bacterial phylogenetic diversity and pH, are strongly consistent with studies performed at the local to continental scales and within a single season^{6,7,81}, which indicates the robustness of our results. Nonetheless, validation of the proposed models needs to be performed by other researchers with more data or an independent dataset, particularly by including samples from under-sampled regions (Extended Data Fig. 1a) and from different seasons (to account for seasonality). Under-sampled regions in our dataset (for example, North Asia) lowered precision of our models for those regions. Unfortunately, there are no published global datasets with comparable sampling protocols that could be directly compared and used for model validation, and we encourage future studies that will make this possible.

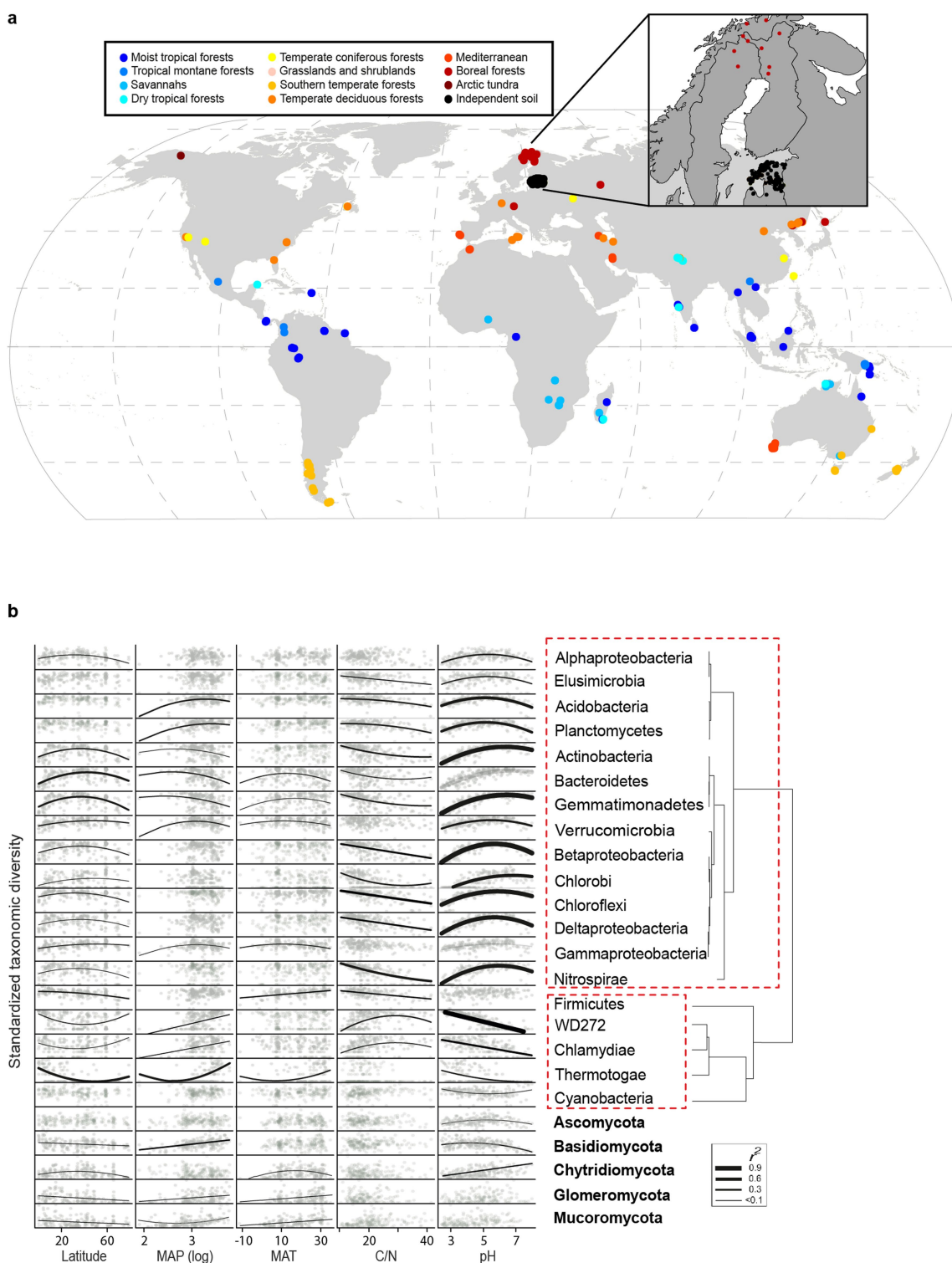
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The pipeline to process metabarcoding samples is available under <http://psbweb05.psb.ugent.be/lotus/>. The pipeline to process shotgun metagenomic samples is available under <https://github.com/hildebra/MATAFILER> and <https://github.com/hildebra/Rarefaction>.

Data availability. All metagenomics and metabarcoding sequences have been deposited in the European Bioinformatics Institute Sequence Read Archive database: Estonian forest and grassland topsoil samples, accession numbers PRJEB24121 (ERP105926); 16S metabarcoding data of global soil samples, accession numbers PRJEB19856 (ERP021922); 18S metabarcoding data of global soil samples, accession numbers PRJEB19855 (ERP021921); Global analysis of soil microbiomes, accession numbers PRJEB18701 (ERP020652). The soil gene catalogue and dataset are available at http://vm-lux.embl.de/~hildebra/Soil_gene_cat/. The Tara Oceans data are available at <http://ocean-microbiome.embl.de/companion.html>. All other data that support the findings of this study are available from the corresponding authors upon request.

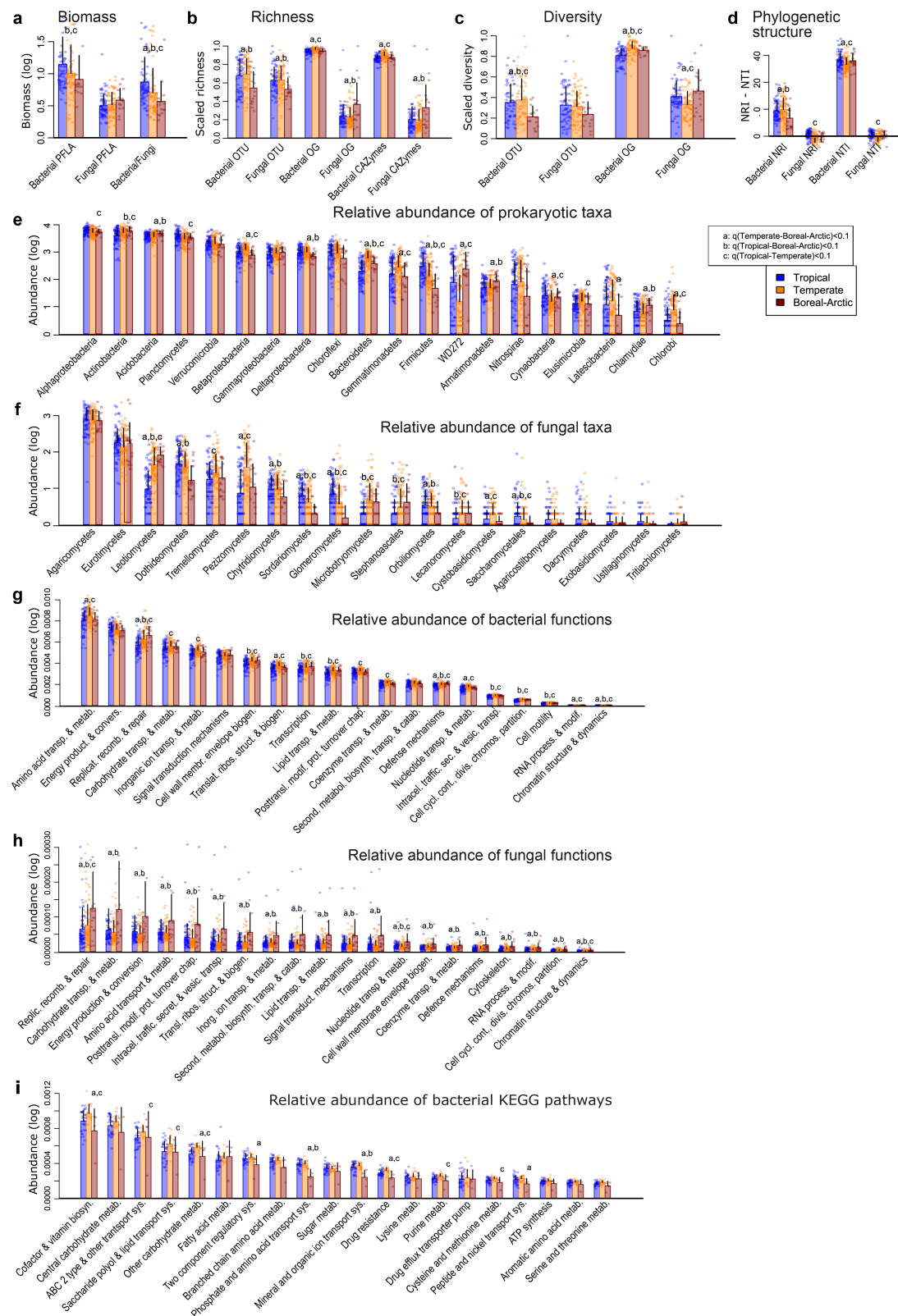
1. Tedersoo, L. et al. Shotgun metagenomes and multiple primer pair–barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *Mycologia* **10**, 1–43 (2015).
2. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
3. Walters, W. et al. Improved bacterial 16S rRNA gene (V4 and V4–5) and fungal internal transcribed spacer marker gene primers for microbial community survey. *mSystems* **1**, e00009–e00015 (2015).
4. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4**, e6372 (2009).

35. Tedersoo, L. et al. Enzymatic activities and stable isotope patterns of ectomycorrhizal fungi in relation to phylogeny and exploration types in an afro-tropical rain forest. *New Phytol.* **195**, 832–843 (2012).
36. Frostegård, Å., Tunlid, A. & Bååth, E. Use and misuse of PLFA measurements in soils. *Soil Biol. Biochem.* **43**, 1621–1625 (2011).
37. van Aarle, I. M. & Olsson, P. A. Fungal lipid accumulation and development of mycelial structures by two arbuscular mycorrhizal fungi. *Appl. Environ. Microbiol.* **69**, 6762–6767 (2003).
38. Frostegård, Å., Tunlid, A. & Bååth, E. Phospholipid fatty acid composition, biomass, and activity of microbial communities from two soil types experimentally exposed to different heavy metals. *Appl. Environ. Microbiol.* **59**, 3605–3617 (1993).
39. Prosser, J. I. Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. *Nat. Rev. Microbiol.* **13**, 439–446 (2015).
40. Salles, J. F., Le Roux, X. & Poly, F. Relating phylogenetic and functional diversity among denitrifiers and quantifying their capacity to predict community functioning. *Front. Microbiol.* **3**, 209 (2012).
41. Trivedi, P. et al. Microbial regulation of the soil carbon cycle: evidence from gene–enzyme relationships. *ISME J.* **10**, 2593–2604 (2016).
42. Puente-Sánchez, F., Aguirre, J. & Parro, V. A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies. *Nucleic Acids Res.* **44**, e40 (2016).
43. Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
44. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
45. Choi, J. et al. fPoxDB: fungal peroxidase database for comparative genomics. *BMC Microbiol.* **14**, 117 (2014).
46. Fawal, N. et al. PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res.* **41**, D441–D444 (2012).
47. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
48. Cantarel, B. L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
49. Cantarel, B. L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS ONE* **7**, e28742 (2012).
50. Cardenas, E. et al. Forest harvesting reduces the soil metagenomic potential for biomass decomposition. *ISME J.* **9**, 2465–2476 (2015).
51. Forslund, K. et al. Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* **23**, 1163–1169 (2013).
52. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).
53. McArthur, A. G. et al. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
54. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
55. Hauswedell, H., Singer, J. & Reinert, K. Lambda: the local aligner for massive biological data. *Bioinformatics* **30**, i349–i355 (2014).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
59. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
60. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
61. Howe, A. C. et al. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl Acad. Sci. USA* **111**, 4904–4909 (2014).
62. Sharon, I. et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res.* **25**, 534–543 (2015).
63. Meyer, F. et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
64. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
65. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
66. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
67. Li, H. Minimap2: fast pairwise alignment for long DNA sequences. Preprint at <https://arxiv.org/abs/1708.01492> (2017).
68. Saary, P., Forslund, K., Bork, P. & Hildebrand, F. RTK: efficient rarefaction analysis of large datasets. *Bioinformatics* **33**, 2594–2595 (2017).
69. Edgar, R. C. UPPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
70. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
71. Guillou, L. et al. The Protist Ribosomal Reference database (PR²): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2012).
72. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
73. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
74. Truong, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
75. Mende, D. R. et al. Assessment of metagenomic assembly using simulated next generation sequencing data. *PLoS ONE* **7**, e31386 (2012).
76. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
77. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
78. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
79. Carini, P. et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2**, 16242 (2016).
80. Žifčáková, L., Větrovský, T., Howe, A. & Baldrian, P. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environ. Microbiol.* **18**, 288–301 (2016).
81. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626–631 (2006).



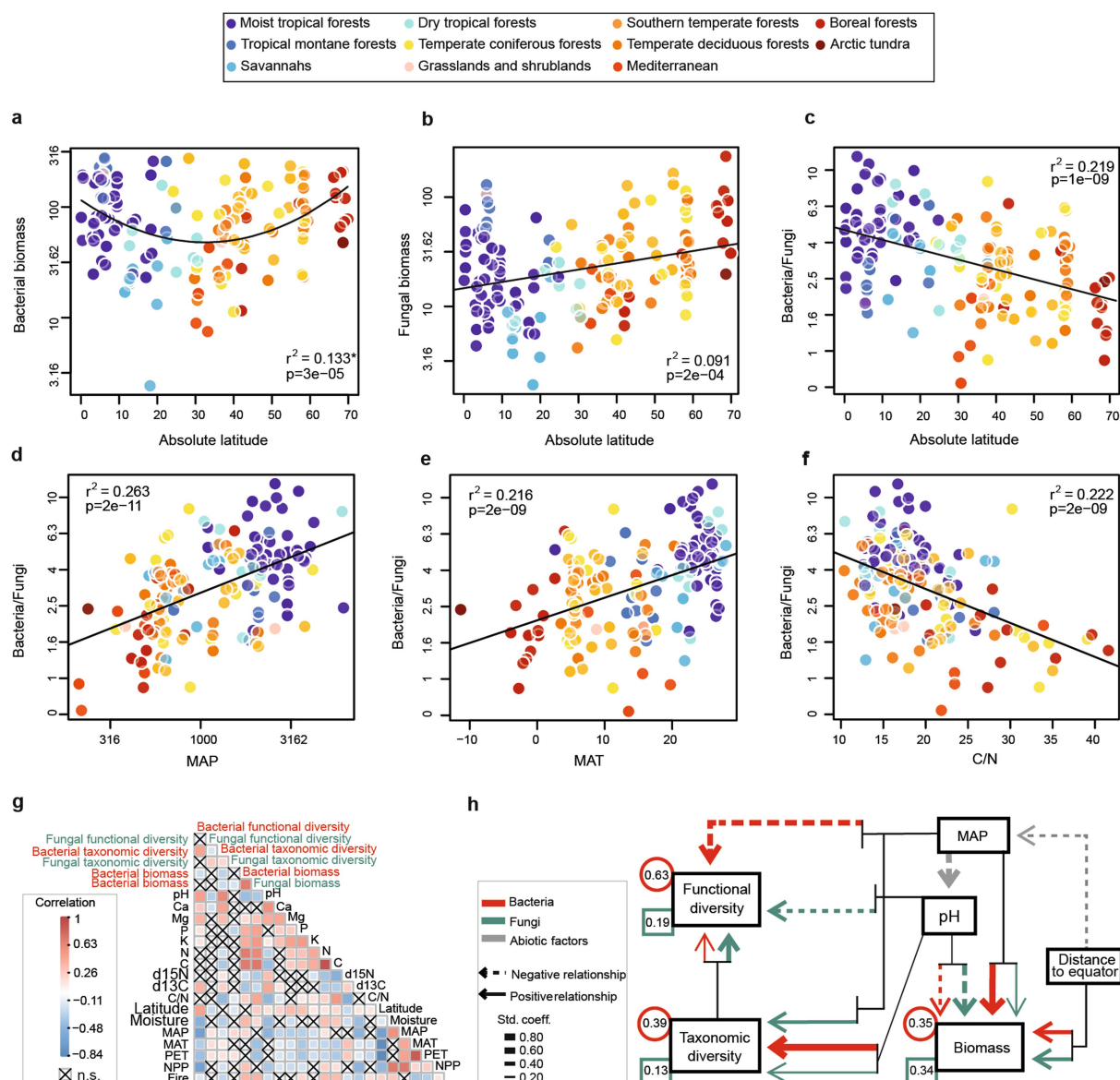
Extended Data Fig. 1 | Distribution of topsoil samples and diversity patterns of phyla. a, A map of samples used for metagenomic and metabarcoding analysis. Colours indicate biomes as shown in the legend. Desert samples were only used in metabarcoding analysis and were excluded in comparative analysis of functional and taxonomic patterns. Black symbols refer to samples from an independent soil dataset (145 topsoil samples; Supplementary Table 1) that were used for validation of

our results. **b,** Relationship between the diversity of major microbial phyla (classes for Proteobacteria) and environmental variables across the global soil samples ($n = 197$ biologically independent samples). Only regression lines for significant relationships after Bonferroni correction are shown. Diversity was measured using Hellinger-transformed matrices on the basis of inverse Simpson index. Latitude, absolute latitude.



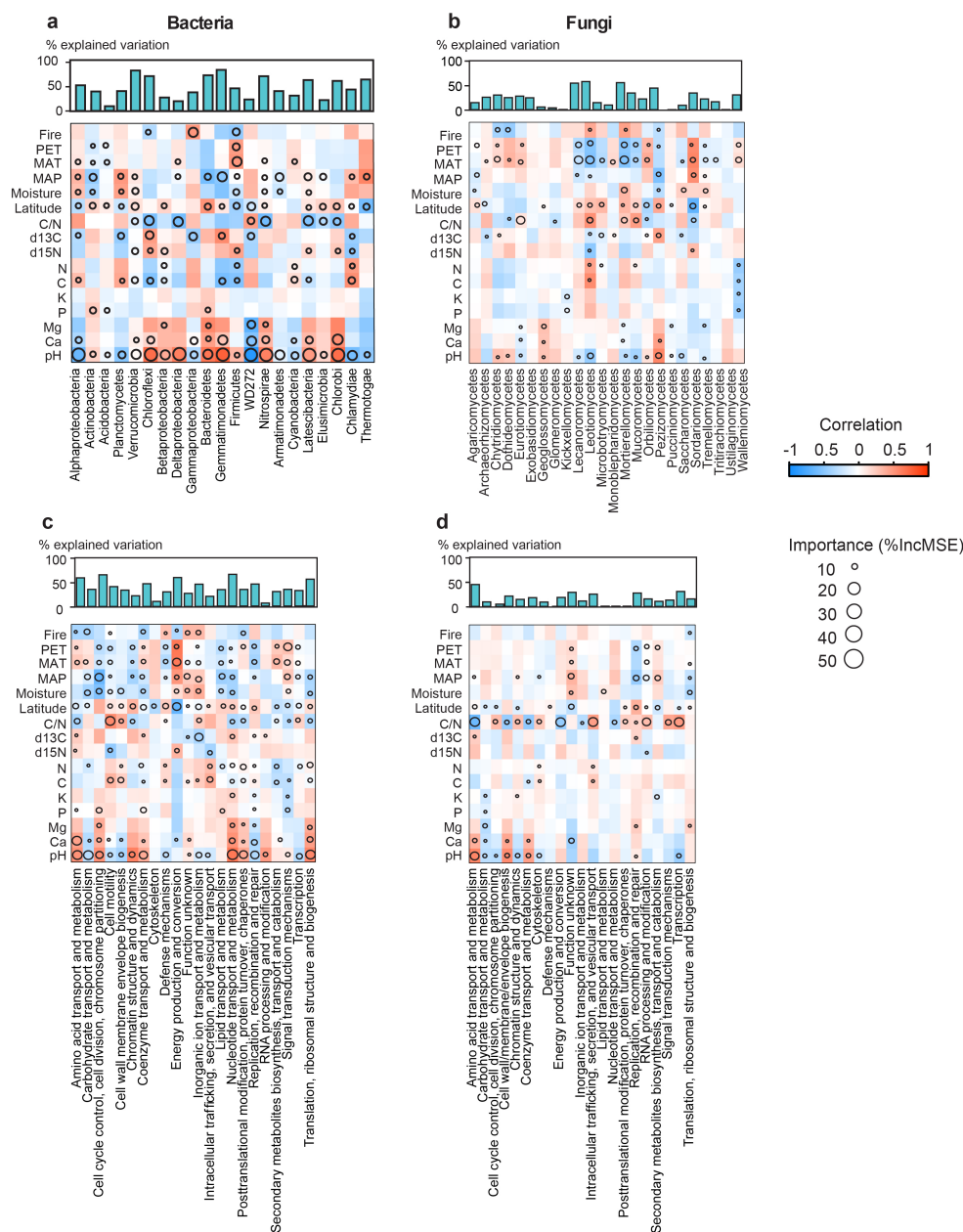
Extended Data Fig. 2 | Contrasting microbial structure and function in major terrestrial biomes. **a–d.** The average total biomass normalized to organic carbon (**a**, $n = 152$ biologically independent samples) as well as richness (**b**), diversity (**c**) and phylogenetic structure including NRI and NTI (**d**) ($n = 188$ biologically independent samples) of fungi and bacteria across samples categorized into major terrestrial biomes, including tropical (moist and dry tropical forests and savannahs), temperate (coniferous and deciduous forests, grasslands and shrublands, and Mediterranean biomes) and boreal–arctic ecosystems. **e–i.** Relative abundance of major phyla ($n = 188$ biologically independent samples) and functional categories

($n = 189$ biologically independent samples) across biomes: bacterial phyla (classes for Proteobacteria) and archaea (**e**); fungal classes (**f**); functional categories of bacteria (**g**); functional categories of fungi (**h**); bacterial KEGG metabolic pathways (**i**). Biomass was measured on the basis of PLFA analysis. Different letters denote significant differences between groups (shown in the legend) at the 0.05 probability level on the basis of Kruskal–Wallis tests corrected for multiple testing. Additional details for these comparisons are presented in Supplementary Table 14. Taxonomic and gene functional diversity indices were calculated on the basis of inverse Simpson index. Data are mean \pm s.d.



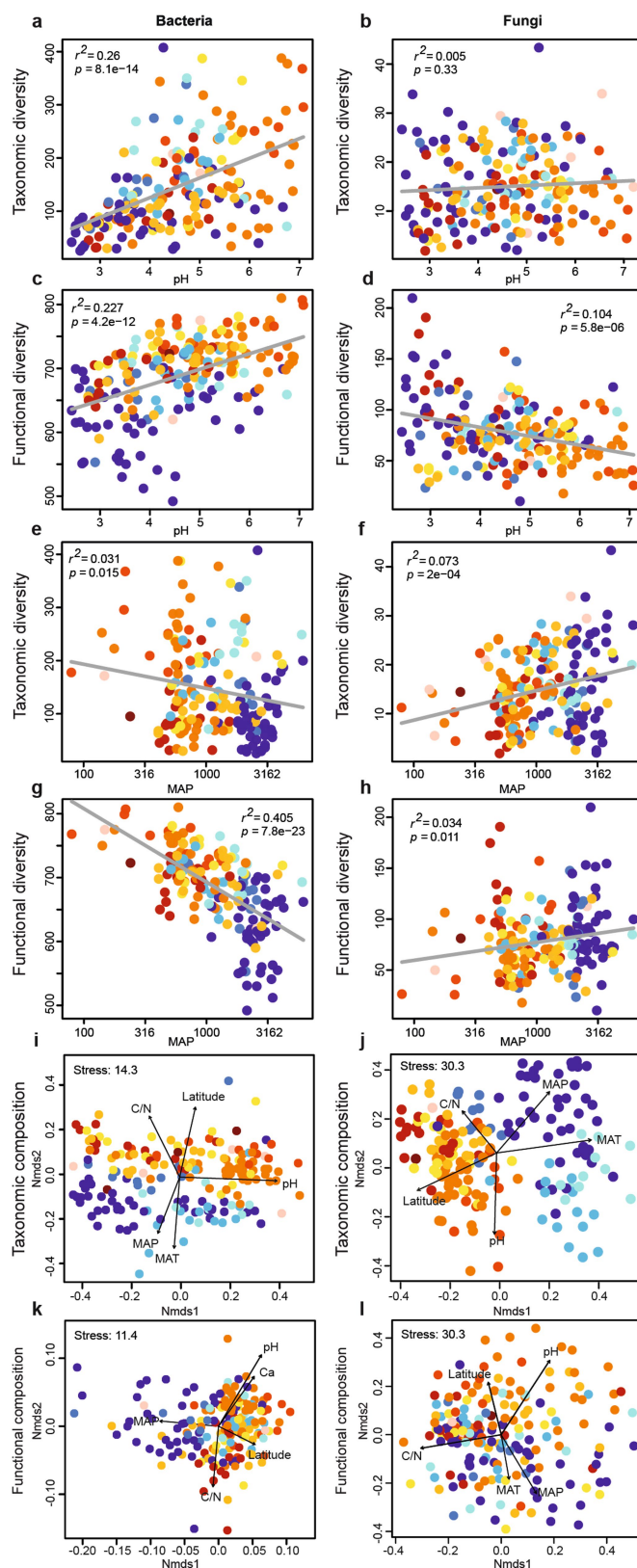
Extended Data Fig. 3 | The significant decrease in the bacterial/fungal biomass ratio with increasing latitude is driven by the joint effect of climate and soil fertility. a, The second order polynomial relationship between absolute latitude and the total biomass of bacteria ($n = 152$ biologically independent samples). **b**, The relationship between absolute latitude and the total biomass of fungi. **c**, The relationship between absolute latitude and the bacterial/fungal biomass ratio. **d–f**, The relationship between bacterial/fungal biomass ratio and MAP, MAT and C/N, as the main correlated environmental variables with bacterial/fungal biomass ratio. Linear regression analysis (Pearson's correlation) was used in **b–f** ($n = 152$ biologically independent samples). **g**, Pairwise Spearman's correlation matrix of biotic and abiotic variables in soil. **h**, Direct and

indirect relationships and directionality between variables determined from best-fitting structural equation model. Determination coefficients (R^2) are given for biomass and diversity factors (see Supplementary Table 5 for more details). Goodness of fit: bacteria, $\chi^2 = 15.37$, degrees of freedom = 11, $P = 0.166$; RMSEA = 0.041, $P_{\text{CLOSE}} = 0.573$, $n = 189$; fungi, $\chi^2 = 7.74$, degrees of freedom = 12, $P = 0.805$; RMSEA = 0.00, $P_{\text{CLOSE}} = 0.970$, $n = 189$). Biomass (nmol g^{-1}) was measured on the basis of PLFA analysis. pH, soil pH representing soil pH and its quadratic term; $\delta^{15}\text{N}$, nitrogen stable isotope signature; $\delta^{13}\text{C}$, carbon stable isotope signature; PET, potential of evapotranspiration; Fire, time from the last fire disturbance; NPP, net primary productivity.



Extended Data Fig. 4 | The environment has a stronger effect on bacterial taxa and functions than on those of fungi. Correlation and best random forest model for major taxonomic (a, b; $n = 188$ biologically independent samples) and functional (c, d; $n = 189$ biologically independent samples) categories of bacteria (a, c) and fungi (b, d) in the global soil samples ($n = 189$ biologically independent samples). a, Relative abundance of major 16S-based bacterial phyla (class for Proteobacteria).

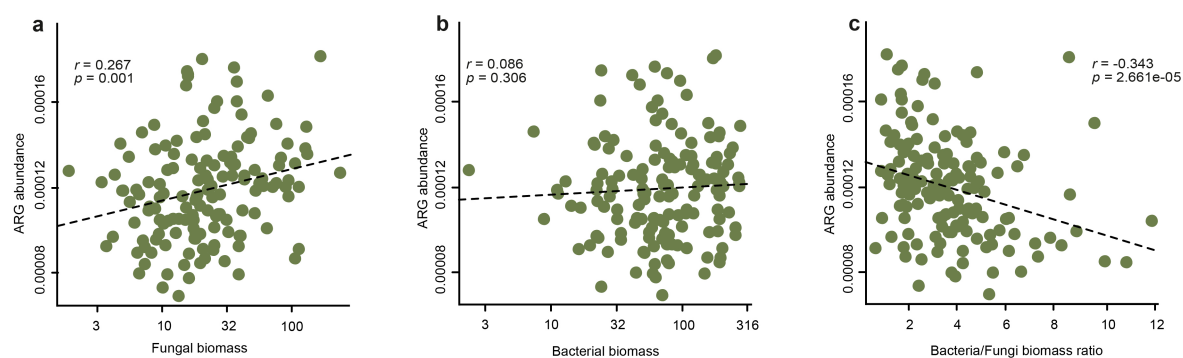
b, Relative abundance of ITS-based fungal classes. c, d, Major orthologous gene categories of bacteria (c) and fungi (d). For variable selection and estimating predictability, the random forest machine-learning algorithm was used. Circle size represents the variable importance (that is, decrease in the prediction accuracy (estimated with out-of-bag cross-validation)) as a result of the permutation of a given variable. Colours represent Spearman correlations. pH, soil pH.



Extended Data Fig. 5 | See next page for caption.

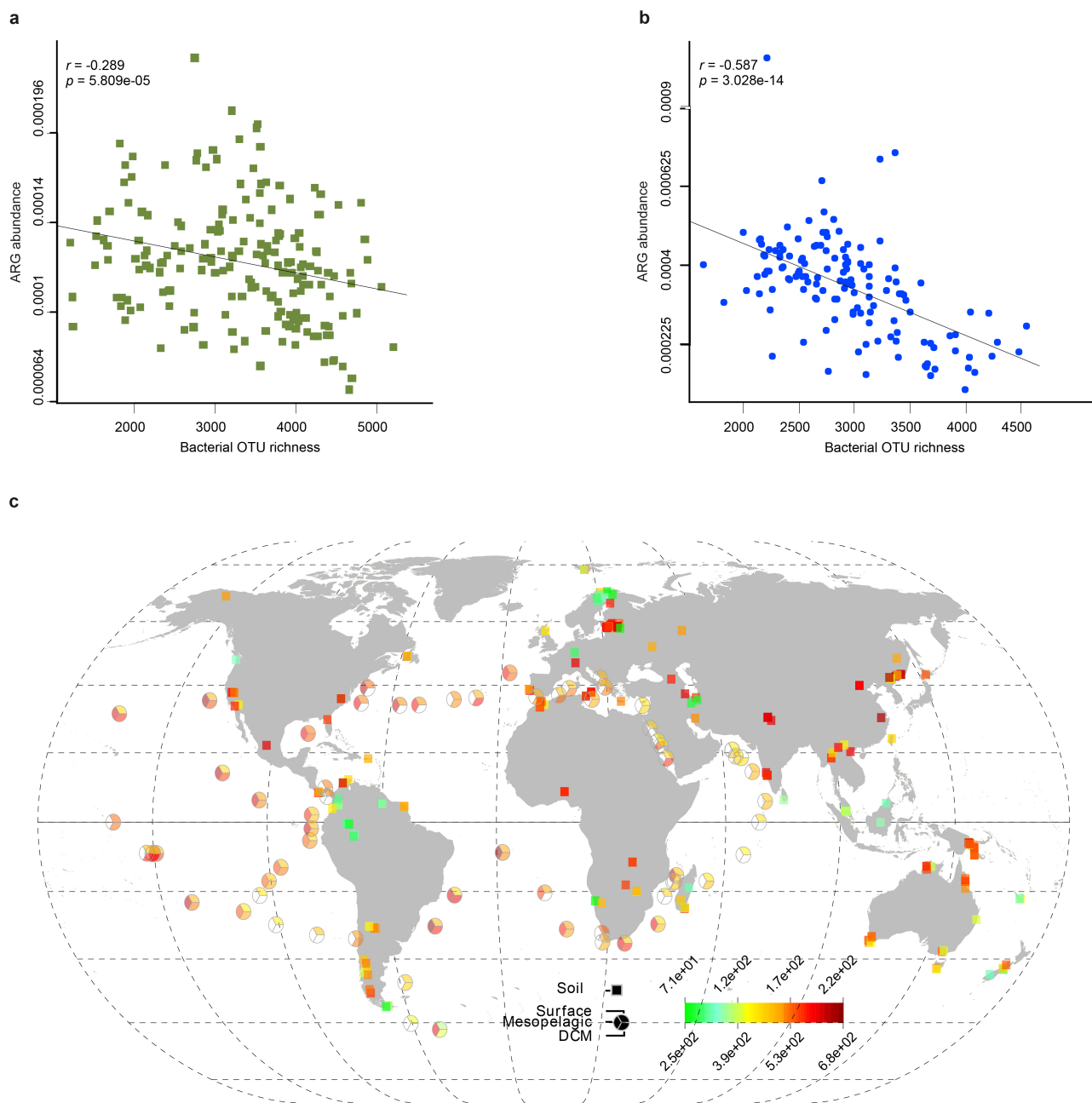
Extended Data Fig. 5 | Niche differentiation between bacteria and fungi is probably related to precipitation and soil pH. Contrasting effect of pH and MAP on bacterial (16S; left column) and fungal (18S; right column) taxonomic ($n = 188$ biologically independent samples) and gene functional ($n = 189$ biologically independent samples) diversity in the global soil samples. **a, b**, Relationship between soil pH and taxonomic diversity of bacteria and fungi. **c, d**, Relationship between soil pH and gene functional diversity of bacteria and fungi. **e, f**, Relationship between MAP and taxonomic diversity of bacteria and fungi. **g, h**, Relationship between MAP and gene functional diversity of bacteria and fungi. Lines represent regression lines of best fit. The choice of degree of polynomial was

determined by a goodness of fit. Colours denote biomes as indicated in the legend. Taxonomic and gene functional diversity indices were calculated on the basis of inverse Simpson index. **i–l**, NMDS plots of trends in taxonomic (16S and 18S datasets) and gene functional composition (orthologous groups from metagenomes) of bacteria and fungi on the basis of Bray–Curtis dissimilarity. **i**, Taxonomic composition of bacteria (16S). **j**, Taxonomic composition of fungi (18S). **k**, Gene functional composition of bacteria. **l**, Gene functional composition of fungi. **i**, Colours denote biomes as indicated in the legend. Vectors are the prominent environmental drivers fitted onto ordination.



Extended Data Fig. 6 | Fungal biomass is significantly related to the relative abundance of ARGs. a, Increase in fungal biomass is related to ARG relative abundance. **b,** Bacterial biomass is unrelated to the relative abundance of ARGs. **c,** ARG relative abundance is inversely correlated

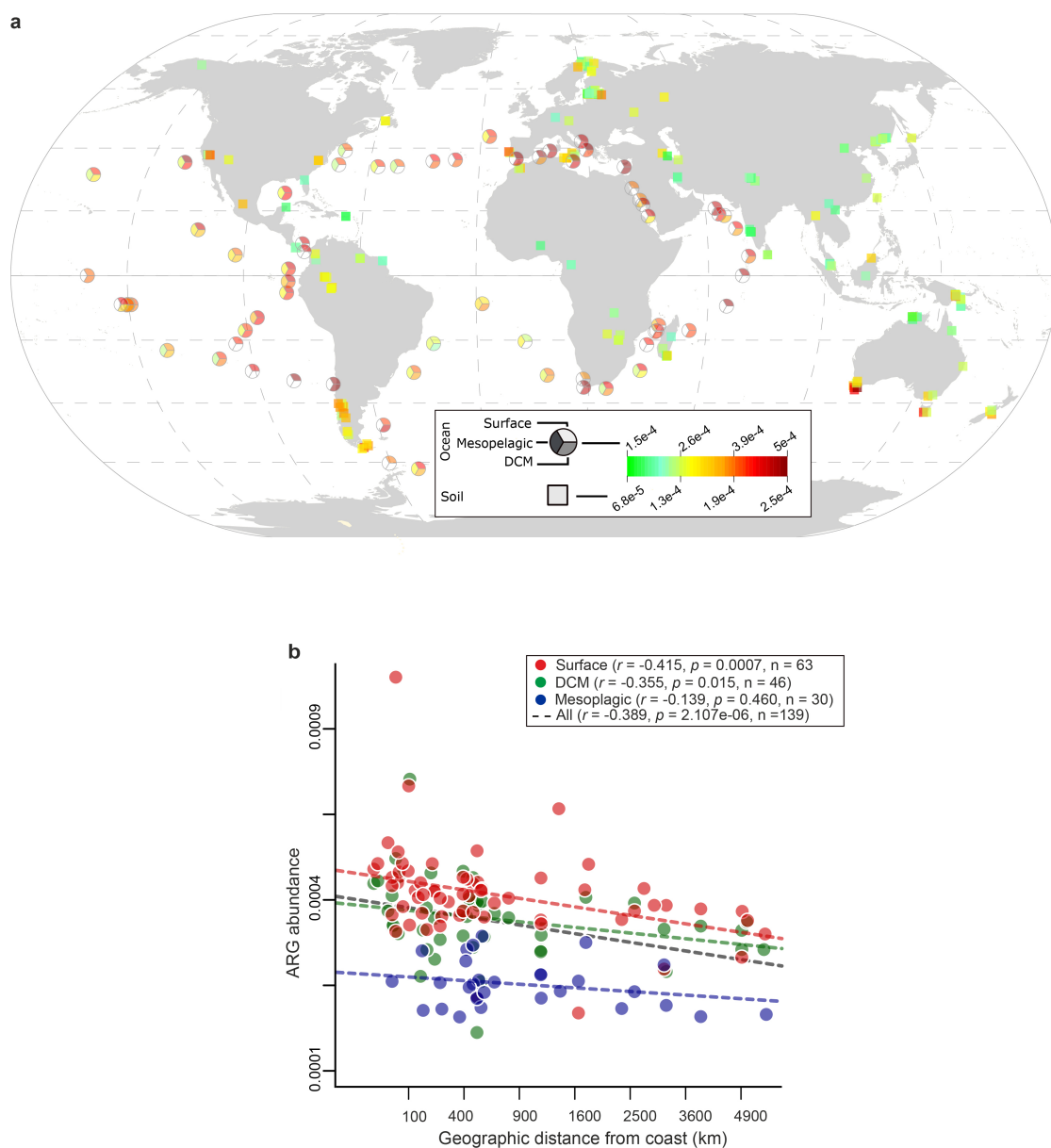
with the bacterial/fungal biomass ratio. Biomass (nmol g^{-1}) was measured on the basis of PLFA analysis. Spearman's correlation was used ($n = 152$ biologically independent samples).



Extended Data Fig. 7 | Topsoil and ocean bacterial phylogenetic diversity is negatively correlated with the abundance of ARGs.

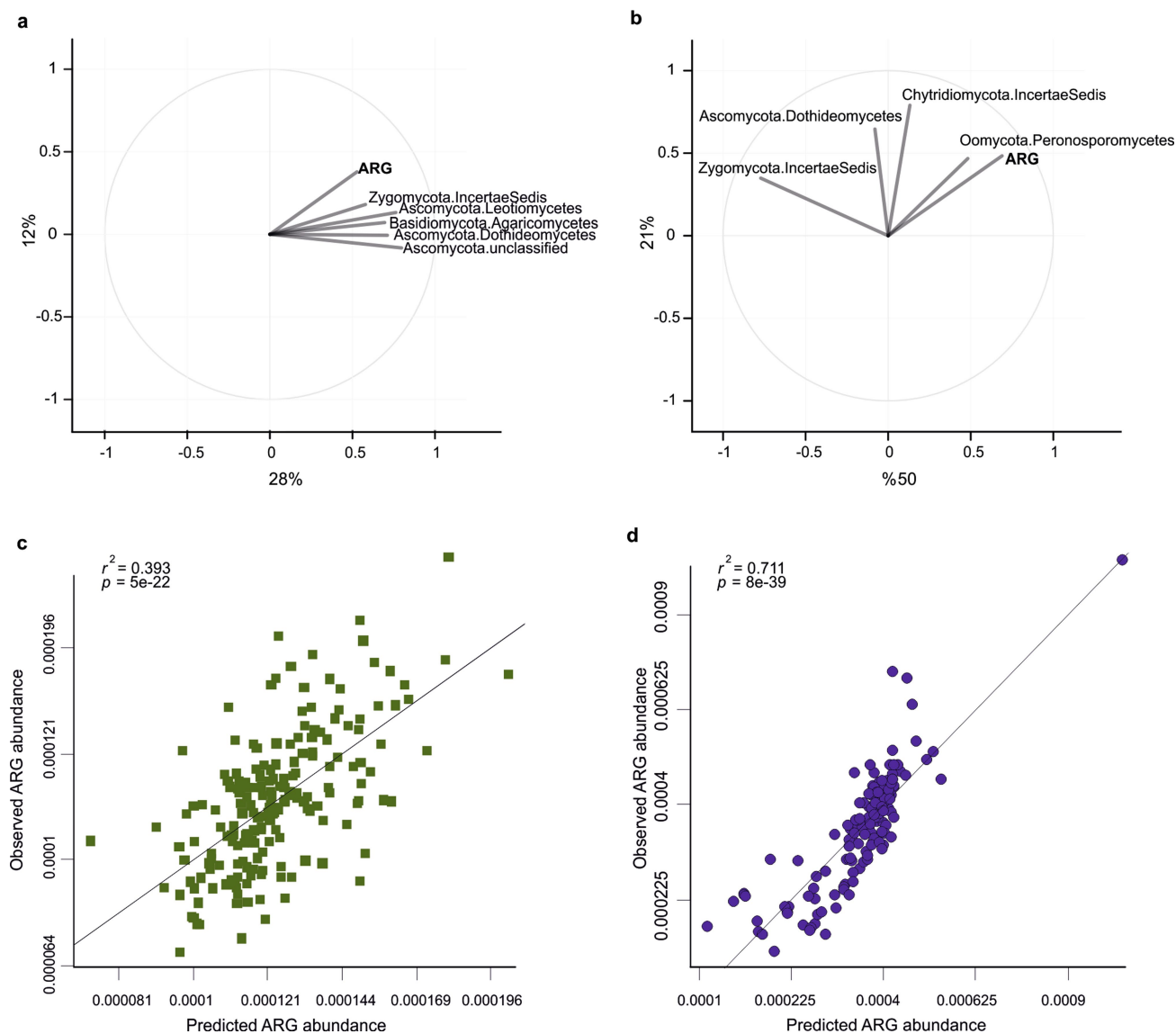
a, b, Spearman's correlation between the relative abundance of ARGs and bacterial phylogenetic diversity (Faith's index) in soil (**a**, $n = 188$ biologically independent samples) and the oceans (**b**, $n = 139$ biologically

independent samples) at the global scale. Similar trends were observed for richness ($r = -0.219$, $P = 0.007$ and $r = -0.659$, $P < 10^{-15}$ in soil and ocean, respectively). **c,** Global map of observed bacterial phylogenetic diversity (Faith's index) at the sampled sites. Note that hotspots of bacterial diversity do not correspond to ARG hotspots (See Extended Data Fig. 8).



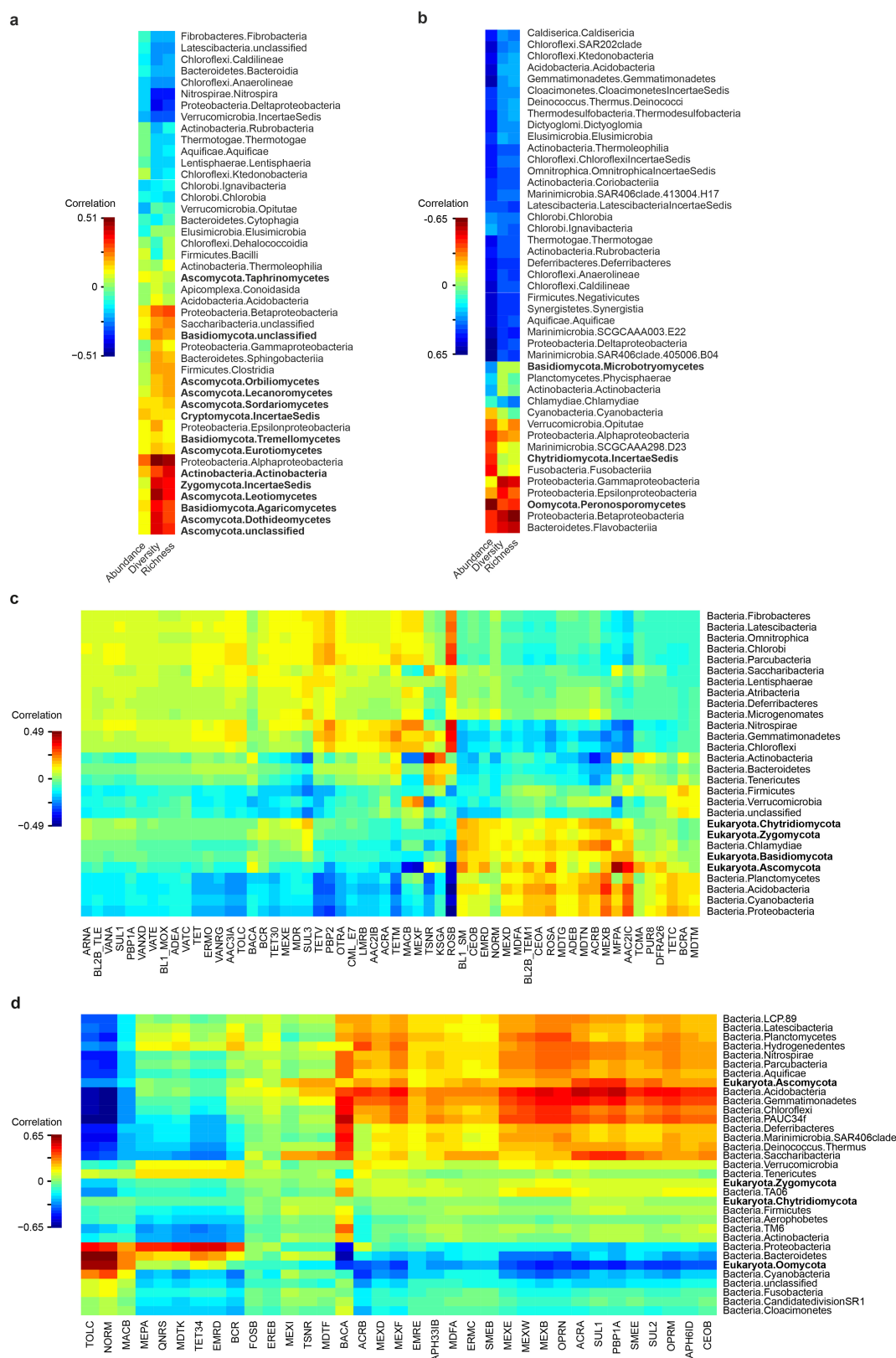
Extended Data Fig. 8 | Relative abundance of ARGs within and between terrestrial and oceanic ecosystems. a, Heat map of the observed relative abundance of ARGs at the global scale. Squares and circles correspond to soil and to ocean samples, respectively. ARG abundance is given on three relative scales for these three datasets. **b,** Relative abundance of ARGs in ocean samples (across depths) declines with the distance from land ($n = 139$ biologically independent samples), a pattern that was significant

at two water depths, including surface (red) and deep chlorophyll maximum (DCM; green), but not at mesopelagic (blue). Spearman's correlation statistics for specified comparisons are given in the legends. Dotted lines display Spearman's correlations across the whole dataset and within the three depth categories, respectively. n , number of biologically independent samples.



Extended Data Fig. 9 | Relative abundance of ARGs in both ocean and topsoil samples can be modelled by the relative abundance of fungi and fungus-like protists. a, b, Correlation circle indicating the relationships among fungal classes and the relative abundance of ARGs as well as the first two PLS components in soil (**a**) and ocean (**b**). Length and direction of vectors indicate the strength and direction of correlations. Percentages show the variation explained by each PLS component. **c, d,** Linear (Pearson) correlations between observed and modelled ARG relative abundance on the basis of the relative abundance of fungal taxa

in soil (**c**) and ocean (**d**). The two principal axes were chosen on the basis of leave-one-out cross-validation (LOOCV) and explained 40% (LOOCV: $R^2 = 0.381$) and 71% (LOOCV: $r^2 = 0.684$) of the variation of the relative abundance of ARGs in soil and the oceans, respectively. Only taxa significantly associated with the relative abundance of ARGs are shown. Cross-validation and LASSO regression confirmed this result. Soil dataset: $r = 0.619$, $RMSE = 10^{-9}$, $n = 189$ biologically independent samples; ocean dataset, $r = 0.832$, $RMSE = 10^{-9}$, $n = 139$ biologically independent samples.



Extended Data Fig. 10 | Fungal classes are among the main taxa associated with the relative abundance, diversity and richness of ARGs in different habitats. a, b, Heat map derived from sPLS analysis showing correlation of total relative abundance, richness and diversity of ARGs to that of the main taxonomic classes in soil (a) and ocean (b) metagenomes (see also the Supplementary Discussion for analogous results in previously published soil (from grasslands, deserts agricultural soils) as well as

human skin and gut samples). For statistical details and significance, see Supplementary Table 8. **c, d,** Heat maps showing correlation of total relative abundance of ARGs to that of the main eukaryotic and prokaryotic taxa in soil (c) and the ocean (d) on the basis of sPLS regression analysis. All matrices were normalized to library size and Hellinger transformation. Fungal and fungal-like classes are shown in bold text. See Supplementary Table 15 for ARG gene letter abbreviations.

Mitochondrial double-stranded RNA triggers antiviral signalling in humans

Ashish Dhir^{1*}, Somdutta Dhir^{1,12}, Lukasz S. Borowski^{2,3,12}, Laura Jimenez⁴, Michael Teitell⁴, Agnès Rötig⁵, Yanick J. Crow^{5,6,7}, Gillian I. Rice⁸, Darragh Duffy^{9,10}, Christelle Tamby⁵, Takayuki Nojima¹, Arnold Munnich⁵, Manuel Schiff⁵, Claudia Ribeiro de Almeida¹, Jan Rehwinkel¹¹, Andrzej Dziembowski^{2,3}, Roman J. Szczesny^{2,3*} & Nicholas J. Proudfoot^{1*}

Mitochondria are descendants of endosymbiotic bacteria and retain essential prokaryotic features such as a compact circular genome. Consequently, in mammals, mitochondrial DNA is subjected to bidirectional transcription that generates overlapping transcripts, which are capable of forming long double-stranded RNA structures^{1,2}. However, to our knowledge, mitochondrial double-stranded RNA has not been previously characterized in vivo. Here we describe the presence of a highly unstable native mitochondrial double-stranded RNA species at single-cell level and identify key roles for the degradosome components mitochondrial RNA helicase SUV3 and polynucleotide phosphorylase PNPase in restricting the levels of mitochondrial double-stranded RNA. Loss of either enzyme results in massive accumulation of mitochondrial double-stranded RNA that escapes into the cytoplasm in a PNPase-dependent manner. This process engages an MDA5-driven antiviral signalling pathway that triggers a type I interferon response. Consistent with these data, patients carrying hypomorphic mutations in the gene *PNPT1*, which encodes PNPase, display mitochondrial double-stranded RNA accumulation coupled with upregulation of interferon-stimulated genes and other markers of immune activation. The localization of PNPase to the mitochondrial inter-membrane space and matrix suggests that it has a dual role in preventing the formation and release of mitochondrial double-stranded RNA into the cytoplasm. This in turn prevents the activation of potent innate immune defence mechanisms that have evolved to protect vertebrates against microbial and viral attack.

Bidirectional transcription of mitochondrial DNA (mtDNA) is an extreme example of convergent transcription in mammalian cells owing to symmetrical synthesis of both the heavy (H) and the light (L) strand encoded RNAs. Notably, nearly the entire L-strand transcript undergoes rapid RNA decay by the RNA degradosome³. This decay process probably prevents the formation of potentially deleterious mitochondrial double-stranded RNA (mtdsRNA). Indeed, among different cellular compartments, mitochondrial RNA (mtRNA) is known to be especially immunogenic⁴. Cellular nucleic acid sensors must discriminate viral nucleic acids from the vast excess of often biochemically indistinguishable cellular RNA and DNA as part of the innate immune response⁵. To achieve this, nucleic acid metabolism is pivotal in suppressing immune responses to self nucleic acids⁶. Recently, numerous pathways have been shown to suppress mtDNA sensing by preventing its escape into the cytoplasm^{7,8}. We sought to determine whether mitochondria are also a source of dsRNA in vivo, and in so doing uncovered a pathway that suppresses the formation of immunostimulatory mtdsRNA.

We used a monoclonal antibody (J2) specific for dsRNA that is widely used to detect viral dsRNA in animals and plants⁹. As shown

previously, HeLa cell infection with the positive-strand RNA virus, encephalomyocarditis virus (EMCV) resulted in strong cytoplasmic dsRNA signals⁹ (Extended Data Fig. 1a, b). Notably, weaker immunofluorescence signals were also observed in uninfected HeLa cells suggesting the existence of cellular dsRNA. To further characterize these cellular immunofluorescence signals, fixed cells were pre-treated with structure-specific RNases. Immunofluorescence signals were sensitive to dsRNA-specific RNase III but not single-stranded RNA (ssRNA)-specific RNase T1 or TURBO DNase confirming the presence of dsRNA at a single-cell level (Extended Data Fig. 1c, d). We then verified the specificity of J2 for dsRNA in vitro using ss- or dsRNA immunoprecipitation experiments (Extended Data Fig. 1e). We next performed J2-immunoprecipitation-based dsRNA sequencing (dsRNA-seq) to identify selected cellular dsRNA (Fig. 1a). Notably, the mitochondrial genome generates nearly all detectable cellular dsRNA with 99% of the reads attributable to the mitochondrial genome (Extended Data Fig. 1f). Furthermore, the RNA sequencing profile showed widespread reads from both the H- and L-strand of mtDNA, implying the presence of intermolecular dsRNA (Fig. 1b). This was confirmed by immunofluorescence as 95% of J2 foci colocalized with mitochondria (Fig. 1c). To rule out potential artefacts caused by the expression of mitochondrial pseudogenes integrated in the nuclear genome, we performed dsRNA staining in mtDNA-depleted HeLa cells obtained by either expressing the herpes simplex virus 1 (HSV-1) protein UL12.5M185 or human uracil-*N*-glycosylase (mUNG1)¹⁰. A lack of J2 signal confirmed that the dsRNA identified in our experiments can be wholly attributed to the mitochondrial genome (Extended Data Fig. 1g).

As dsRNA levels are normally suppressed in the cell, presumably to avoid the induction of an interferon response, we investigated mtdsRNA turnover. Actinomycin D (Act-D) treatment, which inhibits mitochondrial transcription, caused a rapid loss of mtdsRNA, unlike the CDK9 inhibitor DRB which inhibits nuclear RNA polymerase II transcription (Extended Data Fig. 2a). To search for factors involved in mtdsRNA suppression, we focused on the SUV3 and PNPase enzymes (encoded by *SUPV3L1* and *PNPT1* genes, respectively), which are known to be involved in the degradation of L-strand transcripts³. siRNA-mediated depletion of either enzyme resulted in a five- to eightfold increase in dsRNA levels, on the basis of both confocal microscopy (Fig. 1d–f) and flow cytometry (Extended Data Fig. 2b). The same effect was observed with a different set of siRNAs (Extended Data Fig. 2c). Other tested factors involved in the metabolism of mitochondrial nucleic acids had no effect on dsRNA levels (Extended Data Fig. 2d). We next confirmed that this increase in steady-state levels of dsRNA was due to changes in mtdsRNA turnover. Upon Act-D treatment, but not DRB, dsRNA levels in control-siRNA-treated cells

¹Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ²Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland. ³Faculty of Biology, University of Warsaw, Warsaw, Poland. ⁴Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA, USA. ⁵INSERM UMR1163, Institut Imagine, Paris, France. ⁶Paris Descartes University, Sorbonne-Paris-Cité, Institut Imagine, Paris, France. ⁷Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. ⁸Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. ⁹Immunobiology of Dendritic Cells, Institut Pasteur, Paris, France. ¹⁰INSERM U1223, Paris, France. ¹¹MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ¹²These authors contributed equally: Somdutta Dhir, Lukasz S. Borowski. *e-mail: ashish.dhir@path.ox.ac.uk; rszczesny@ibb.waw.pl; nicholas.proudfoot@path.ox.ac.uk

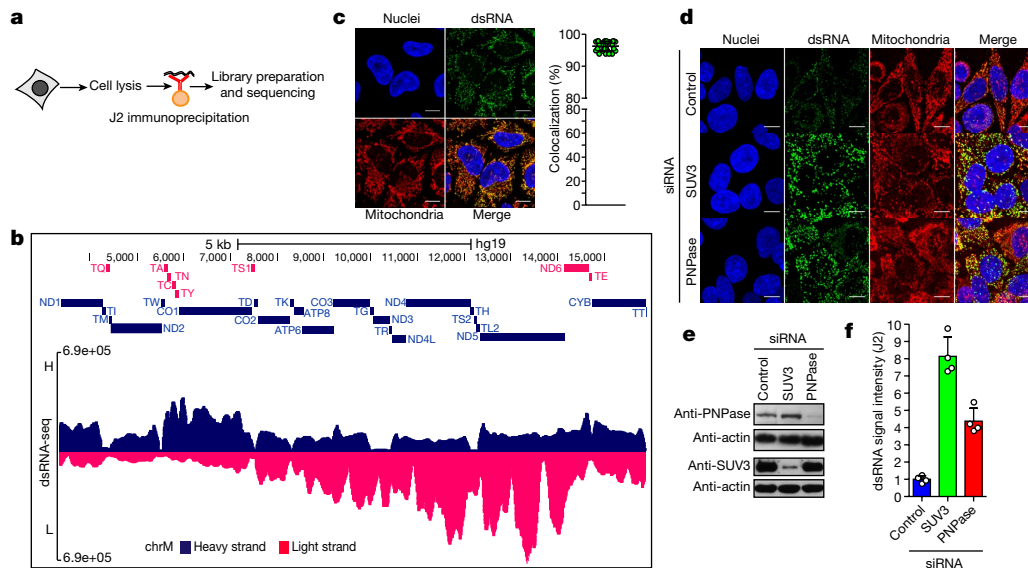


Fig. 1 | Mitochondria form dsRNA that is suppressed by the RNA degradosome. **a**, dsRNA-seq experimental approach. **b**, dsRNA-seq reads across the mitochondrial genome spanning protein coding region in untreated HeLa cells. H-strand genes are shown as blue bars and L-strand as red bars. Short tRNA genes are denoted with T as the first letter. Data are representative of two experiments. **c**, Immunostaining of dsRNA in HeLa cells with anti-dsRNA (J2) antibody. Mitochondria and nuclei are stained with MitoTracker Deep Red and Hoechst, respectively. Scale bars,

10 μ m. Graphs quantify co-localization of dsRNA foci with mitochondria. Data are mean \pm s.d. from 29 cells. **d**, Anti-dsRNA (J2) staining in HeLa cells depleted for PNPase or SUV3 by siRNA as in **c**. Different imaging settings were applied in panel **c** and **d** so that the J2 intensity of control cells varies. **e**, Western blot showing PNPase or SUV3 depletion. Blots are representative of four experiments. **f**, Quantification of dsRNA levels in PNPase- or SUV3-depleted cells. Data are mean \pm s.d. from four experiments. For gel source data, see Supplementary Fig. 1.

were rapidly turned over (half-life of 30 min) whereas dsRNA levels were relatively stable for up to 3 h in either SUV3- or PNPase-depleted cells (Extended Data Fig. 2e).

To further understand the mechanism of dsRNA turnover by SUV3 and PNPase, we used their catalytic mutants. Overexpression of a SUV3 transgene carrying an inactivating mutation (G207V) in the Walker A motif of the helicase in HEK 293 cells acted as a dominant-negative protein¹¹ resulting in accumulation of dsRNA (Extended Data Fig. 3a). Furthermore, northern-blot analysis of J2-immunoprecipitated dsRNA isolated from this dominant negative mutant showed the accumulation of long dsRNA species (approximately 1–6 kb) mapping over the entire mitochondrial genome (Extended Data Fig. 3b). Both RNA import and RNA turnover functions have been ascribed to PNPase^{3,12}. Therefore, an R445E/R446E mutant of PNPase, which lacks exonuclease activity without affecting RNA import, was used^{3,12} (Extended Data Fig. 4a). dsRNA levels accumulating upon PNPase depletion were suppressed by overexpression of siRNA-resistant PNPase but not the R445E/R446E mutant in HeLa cells (Extended Data Fig. 4b–d) and HEK 293 cells (data not shown). Overall, these results implicate the unwinding activity of SUV3 and the exonuclease activity of PNPase in dsRNA turnover. Consistently, J2-immunoprecipitation dsRNA-seq of SUV3- and PNPase-depleted HeLa cells showed substantial accumulation of mtDNA as compared to control siRNA, which was highly reproducible (Extended Data Fig. 5a, b).

As long dsRNA is a hallmark of viral replication that triggers a type I interferon response, *IFNB1* induction was tested in various knock-downs of mitochondrial RNA processing factors. Quantitative PCR with reverse transcription (RT-qPCR) analysis revealed an approximately 90-fold induction of *IFNB1* mRNA upon depletion of PNPase but not upon depletion of SUV3 or MRPP1 (Fig. 2a). Consistently, gene-expression profiling revealed activation of interferon-stimulated genes (ISGs) such as genes with direct antiviral activity (for example, *IFI44*, *IFIT1*), cytoplasmic RNA sensors *DDX58* and *IFIH1* (encoding RIG-I and MDA5, respectively) and the transcription factor *IRF7* that positively reinforces the antiviral response (Extended Data Fig. 6a). The observation that mtDNA activated an interferon response upon depletion of PNPase, but not upon depletion of SUV3, suggested that

SUV3-restricted mtDNA is either non-immunogenic or somehow concealed from cytosolic dsRNA sensors. We therefore isolated mtDNA from mitochondria depleted of SUV3 or PNPase using a magnetic-activated cell sorting (MACS) approach¹³ and transfected it into HeLa cells to induce *IFNB1* mRNA (Fig. 2b). Notably, mtDNA extracted from either condition triggered a similar *IFNB1* induction, which was RNase III sensitive (Fig. 2b). The latter finding confirms that the interferon induction is triggered by mtDNA and not by mtDNA. The experiment also excludes the possibility that SUV3-dependent mtDNA is non-immunogenic, and led us to explore dsRNA localization. Transmission electron microscopy with immunogold labelling using J2 demonstrated mitochondrial localization of dsRNA in control siRNA samples and substantial accumulation in SUV3-depleted cells (Fig. 2c). By contrast, in PNPase-depleted cells, J2 staining displayed both a mitochondrial and cytoplasmic distribution, indicating the release of mtDNA into the cytoplasm (Fig. 2c). Consistently, enhanced mitochondrial outer membrane permeabilization of PNPase-depleted cells using ABT-737 (Bcl-2 inhibitor) resulted in an approximately threefold greater induction of *IFNB1* mRNA (Extended Data Fig. 6b). Lack of an interferon response in SUV3-depleted cells with or without ABT-737 treatment suggested that mtDNA remains restricted to mitochondria (Extended Data Fig. 6b). We confirmed ABT-737-mediated mitochondrial outer membrane permeabilization through release of intermembrane-space-localized protein cytochrome *c* into the cytoplasm (Extended Data Fig. 6c).

We wished to extend our results on PNPase-restricted mtDNA in HeLa cells to an animal gene knockout model. We therefore used the hepatocyte-specific *Pnpt1*^{HepKO} (hereafter HepKO) mouse that has a liver-specific knockout of PNPase as previously described¹². We consistently observed an accumulation of dsRNA in HepKO liver sections versus controls (Fig. 2d). Notably, HepKO cells showed a gradual loss of mtDNA over time, suggesting an adaptive response to interferon activation (M.T., unpublished results), and probably accounting for the heterogeneous increase in dsRNA levels (Fig. 2d, right). However, differential gene-expression analysis showed upregulation of *Ifnb1* and numerous ISGs such as *Ifi44*, *Ifit1*, and *Cxcl10* in HepKO mice (Fig. 2e, Extended Data Fig. 6d). These results are consistent with activation of a

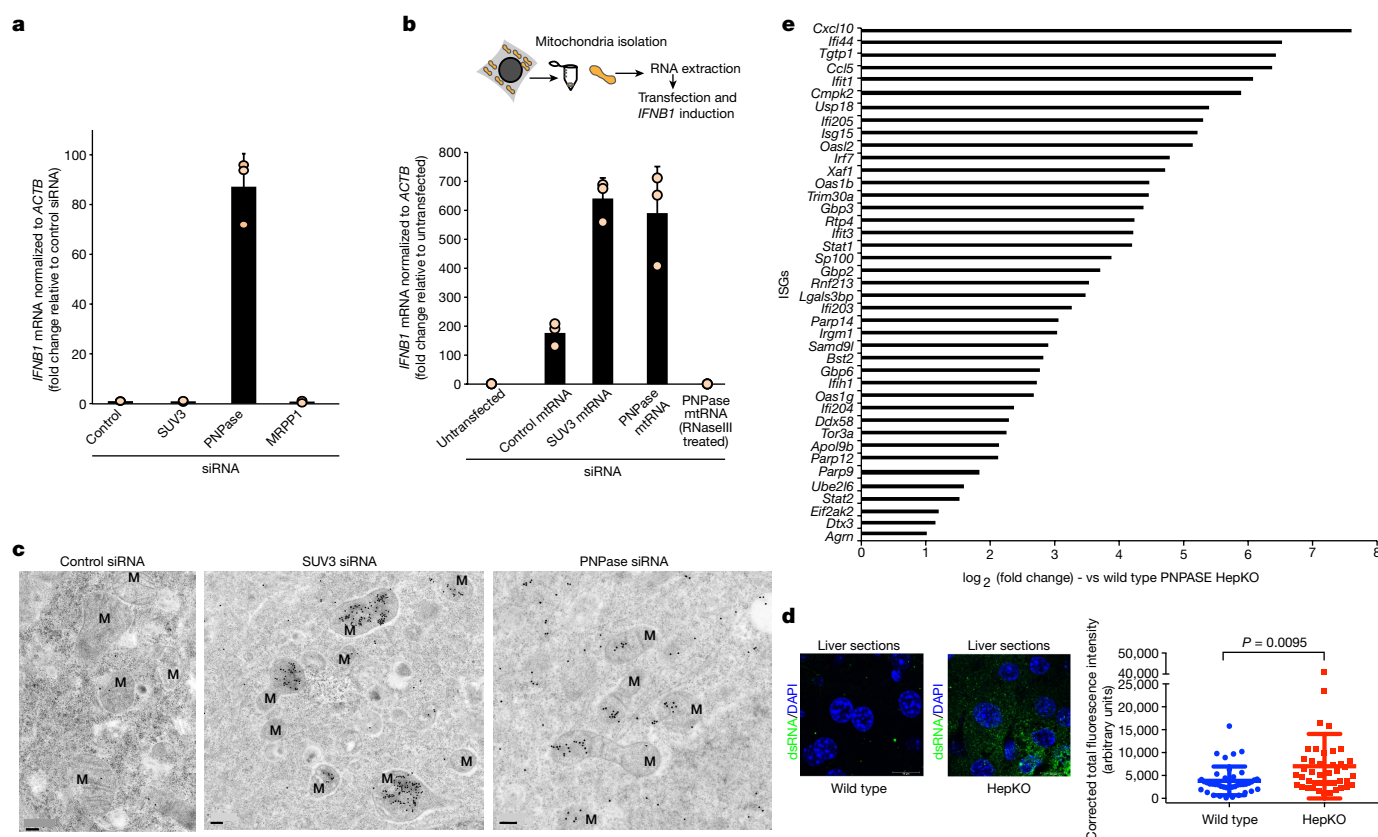


Fig. 2 | PNPase suppresses a mtdsRNA-mediated type I interferon response. **a**, RT-qPCR analysis of *IFNB1* mRNA in HeLa cells treated with the indicated siRNAs. Data are mean \pm s.d. from three independent experiments. **b**, Top, schematic of MACS strategy to purify mtRNA. Bottom, RT-qPCR analysis of *IFNB1* mRNA in HeLa cells transfected with mtRNA (using MACS) as indicated. Data are mean \pm s.d. from three independent experiments. **c**, Transmission electron microscopy images of immunogold labelled dsRNA (J2) in cryofixed HeLa cells treated with the indicated siRNAs. Images are representative of two

experiments. M, mitochondria. Scale bars, 0.2 μ m. **d**, Left, fluorescence immunohistochemistry staining of dsRNA (J2) in liver sections from wild-type and HepKO mice. Nuclei are stained with DAPI. Right, dsRNA quantification. Data are mean \pm s.e.m. from 41 (wild-type) and 42 (HepKO) randomly sampled regions in two liver sections measured. The *P* value is from a two-sided unpaired *t*-test with Welch's correction. **e**, log₂ fold expression change of ISGs in wild-type versus HepKO mice. The ISG list is based on previously published work⁸.

type I interferon response upon loss of PNPase, and support our HeLa cell siRNA-depletion data in murine primary cells.

The importance of PNPase in the restriction of mtdsRNA led us to examine primary fibroblast cells from four different patients carrying biallelic hypomorphic mutations in the *PNPT1* gene identified by exome sequencing and clinical manifestation (Table 1, Extended Data Table 1). These *PNPT1* mutations led to decreased PNPase protein levels in fibroblasts of patients 2, 3 and to some extent 4¹⁴. However, the homozygous active site mutation (R136H) recorded in patient 1 did not¹⁵ (Fig. 3a). Fibroblasts from all four patients demonstrated an accumulation of dsRNA (J2 signal) that was not observed in control cells (Fig. 3b). Moreover, this dsRNA colocalized with mitochondria (Fig. 3b, inset).

The identity of this mtdsRNA signal was further established by RT-qPCR analysis showing the presence of mtdsRNA

(RNase III-sensitive) in pure cytosolic fractions from cells of patients 1 and 2 (Fig. 3c). Also, for three of the patients (no sample was available from patient 1), we recorded an upregulation of ISGs in peripheral blood (Fig. 3d). Furthermore, in patient 2, IFN α protein measured using a digital enzyme-linked immunosorbent assay (ELISA) was increased (603 fg l⁻¹) in cerebrospinal fluid (CSF), which is equivalent to levels observed in certain cases of viral meningitis¹⁶ (Table 1). Patient 2 also showed abnormally high levels of neopterin in the CSF (101 nmol l⁻¹), consistent with a hyperactivated immune response¹⁷ (Table 1). Overall, our analysis of patients harbouring hypomorphic mutations in *PNPT1* clearly underlines the importance of preventing cytosolic sensing of mtdsRNA.

We sought to determine the mechanism of interferon activation by mtdsRNA in the context of PNPase deficiency. We tested the involvement of the RNA sensors RIG-I, MDA5 and TLR3 in this process.

Table 1 | Summarized data from patients with *PNPT1* mutations

Patient	<i>PNPT1</i> mutation	Amino acid change	Effect on PNPase	Outcome	IFN α levels in CSF (fg l ⁻¹) ^a	Neopterin levels in CSF ^b
1	G407A Homozygous	R136H	Abolishes active site	Died aged 2 years	NA	NA
2	T208C G2137T Heterozygous	S70P D713Y	Reduced protein level	Alive at age 1 year	603	101
3	G1495C G1519T Heterozygous	G499R A507S	Reduced protein level	Alive at age 7 years	ND	ND
4	A1160G Homozygous ¹⁴	Q387R	Trimerization defective; reduced protein level	Alive at age 13 years	ND	ND

NA, not available; ND, not determined. See also Extended Data Table 1.

^aIFN α normal range is <1 fg l⁻¹.

^bNeopterin normal range is 8–43 nmol l⁻¹.

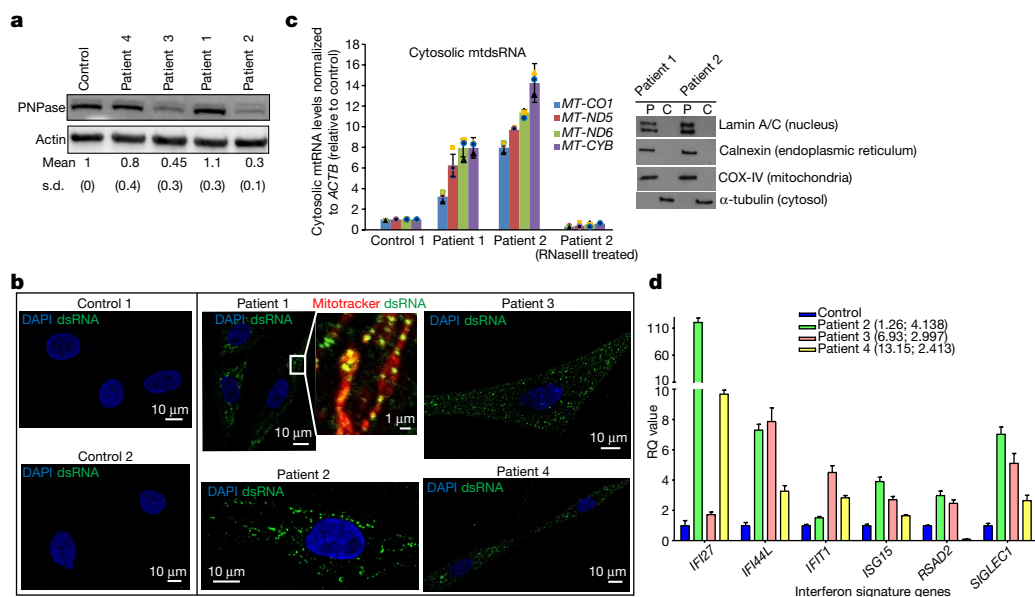


Fig. 3 | Pathological *PNPT1* mutations result in mtDNA accumulation and activation of ISGs. **a**, PNPase western blot in fibroblasts from four patients with mutations in *PNPT1* and a control. Quantification is shown as mean and s.d. from four experiments. **b**, Immunostaining of dsRNA (J2) in fibroblast cell lines from patients with *PNPT1* mutations and controls. Mitochondria are stained with MitoTracker Red CMXRos and nuclei with DAPI (blue). Scale bars, 10 μ m (main images) and 1 μ m (expanded view). Images are representative of two experiments. **c**, Left, RT-qPCR analysis of cytosolic mtDNA (four loci) in cells with *PNPT1* mutations versus

control cells. Data are mean \pm s.d. from three independent experiments. Right, fraction purity as shown by western blots. Blots are representative of two experiments. P, pellet; C, cytosolic fractions. **d**, RT-qPCR analysis of six ISGs in whole blood from patients 2, 3 and 4. Ages when tested in decimalized years and interferon score are shown in brackets. The data plotted is relative quantification (RQ) values for each patient, with the error bars representing RQ_{min} and RQ_{max}. Data are from combined 29 control samples (blue bar) and 3 individual patient samples measured in triplicate. For gel source data, see Supplementary Fig. 1.

In PNPase-depleted HeLa cells, siRNA knockdown of MDA5 or (to a lesser extent) RIG-I abrogated the interferon response, but knockdown of TLR3 did not (Fig. 4a). These data implicate MDA5 as the primary sensor of mtDNA. MDA5 signals via the mitochondrial antiviral signalling protein (MAVS) to induce type I interferons so MAVS knockdown also abrogated *IFNB1* induction (Fig. 4a, Extended Data Fig. 6e). We further confirmed these results by transfecting mtDNA isolated from PNPase-depleted HeLa cells (as in Fig. 2b) into RIG-I-deficient (*Ddx58*^{-/-}) or MDA5-deficient (*Ifih1*^{-/-}) murine embryonic fibroblasts (MEFs) (Fig. 4b). *Ifih1*^{-/-} cells, but not *Ddx58*^{-/-} cells, failed to upregulate mRNA levels of the ISG *Ifit1* in response to mtDNA. This strongly suggests that the mtDNA-induced interferon response is mediated through the MDA5-MAVS axis. The possibility that mtDNA release into cytoplasm involves Bcl2-associated X protein (Bax)-Bcl-2 homologous antagonist/killer (Bak), as in the case of mtDNA^{7,18} was also investigated. Notably, depletion of Bax-Bak prevented *IFNB1* mRNA induction after PNPase depletion, suggesting that mtDNA release depends on Bax-Bak pores (Fig. 4c). As a final test for the escape of mtDNA into the cytoplasm after PNPase depletion, we tested for dsRNA editing by the adenosine deaminase ADAR1¹⁹. Notably, 16 mitochondrial RNA editing sites (including six adenosines to inosines) were observed within the RNA sequencing data of PNPase depleted cells, whereas only one was observed in the SUV3-depleted sample (Extended Data Fig. 7a, b). Concurrent depletion of ADAR1 and PNPase enhanced the observed interferon response by 1.5-fold, suggesting that ADAR1 acts as a feedback suppressor of the innate immune response, which is activated by mtDNA (Extended Data Fig. 7c, d). Overall, these mechanistic data on mtDNA formation, export and engagement with dsRNA sensors can be summarized in a model in which the escape of mtDNA into the cytoplasm triggers an ‘inappropriate’ type I interferon response (Fig. 4d).

Our findings highlight an important function of PNPase which is underscored by its embryonic lethality in knockout mice and its identification as an essential fitness gene in CRISPR screens of human cell lines^{12,20}. We considered it plausible that dysregulation of such an important pathway might induce an innate immune response,

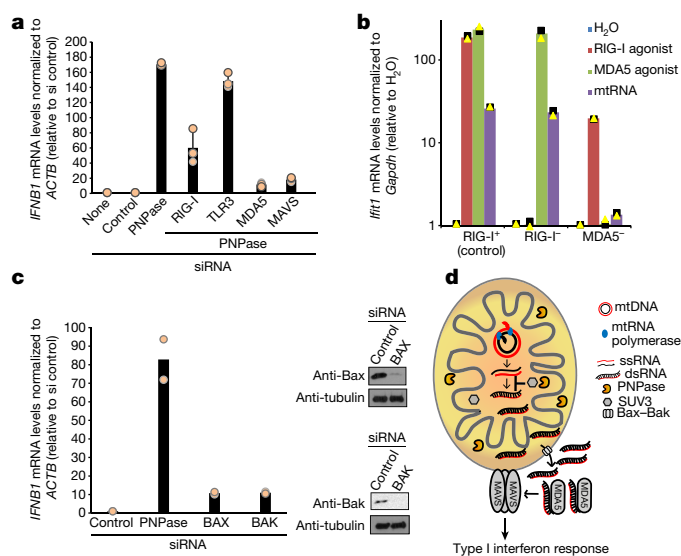


Fig. 4 | MDA5 is the primary sensor of cytosolic mtDNA released in a Bax-Bak dependent fashion. **a**, RT-qPCR analysis of *IFNB1* expression in HeLa cells transfected with indicated siRNAs. Data are mean \pm s.d. from three independent experiments. **b**, RT-qPCR analysis of *Ifit1* expression in *Ddx58*^{+/-} (control, RIG-I⁺), *Ddx58*^{-/-} (RIG-I⁻) and *Ifih1*^{-/-} (MDA5⁻) immortalized MEFs transfected with mtDNA, RIG-I (ppp-IVT-RNA^{99nt}) or MDA5 (CIP-EMCV RNA) specific agonists. Data are the mean from two independent experiments. Values are plotted on a logarithmic scale. **c**, Left, RT-qPCR analysis of *IFNB1* mRNA in HeLa cells treated with the indicated siRNAs. Data are mean from two independent experiments. Right, western blot of siRNA depletion efficiency. Blots are representative of two experiments. For gel source data, see Supplementary Fig. 1. **d**, Model of mtDNA release and Type I interferon response. PNPase restricts mtDNA in matrix (together with SUV3) and IMS. MDA5 acts as the primary mtDNA sensor transducing an interferon response through the MAVS signalling pathway.

consistent with a disease class referred to as the type I interferonopathies¹⁶. Indeed, we show that biallelic hypomorphic mutations in *PNPT1* cause mtdsRNA accumulation and immune activation. We suggest that mtdsRNA is a key mitochondrial-derived agonist of the innate immune system, a role until now mainly attributed to mtDNA²¹. Of note, genetic variants in MDA5 have been implicated in a number of human pathologies, both monogenic, and complex^{22–27}. It is plausible that mtdsRNA mislocalization into the cytosol triggers an innate immune response upon viral infection, as dsRNA accumulation is detectable upon viral infection of mammalian cells⁹. Notably, dsRNA accumulation upon EMCV infection in HeLa cells partially colocalizes with mitochondria (Extended Data Fig. 8). It is possible that cellular mtdsRNA accumulation and the escape of mtdsRNA into the cytoplasm upon viral infection prime an antiviral response, as shown for mtDNA⁸. Overall, our results demonstrate a fundamental role of mitochondrial RNA processing in preventing the accumulation of deleterious self nucleic acid such as dsRNA that would otherwise trigger innate immunity.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0363-0>.

Received: 21 February 2017; Accepted: 6 June 2018;

Published online 25 July 2018.

- Aloni, Y. & Attardi, G. Symmetrical in vivo transcription of mitochondrial DNA in HeLa cells. *Proc. Natl Acad. Sci. USA* **68**, 1757–1761 (1971).
- Young, P. G. & Attardi, G. Characterization of double-stranded RNA from HeLa cell mitochondria. *Biochem. Biophys. Res. Commun.* **65**, 1201–1207 (1975).
- Borowski, L. S., Dziembowski, A., Hejnowicz, M. S., Stepień, P. P. & Szczesny, R. J. Human mitochondrial RNA decay mediated by PNPase-hSuv3 complex takes place in distinct foci. *Nucleic Acids Res.* **41**, 1223–1240 (2013).
- Karikó, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 (2005).
- Schlee, M. & Hartmann, G. Discriminating self from non-self in nucleic acid sensing. *Nat. Rev. Immunol.* **16**, 566–580 (2016).
- McGlasson, S., Jury, A., Jackson, A. & Hunt, D. Type I interferon dysregulation and neurological disease. *Nat. Rev. Neurol.* **11**, 515–523 (2015).
- Rongvaux, A. et al. Apoptotic caspases prevent the induction of type I interferons by mitochondrial DNA. *Cell* **159**, 1563–1577 (2014).
- West, A. P. et al. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–557 (2015).
- Weber, F., Wagner, V., Rasmussen, S. B., Hartmann, R. & Paludan, S. R. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J. Virol.* **80**, 5059–5064 (2006).
- Spadafora, D., Kozhukhar, N., Chouljenko, V. N., Kousoulas, K. G. & Alexeyev, M. F. Methods for efficient elimination of mitochondrial DNA from cultured cells. *PLoS ONE* **11**, e0154684 (2016).
- Szczesny, R. J. et al. Human mitochondrial RNA turnover caught in flagrante: involvement of hSuv3p helicase in RNA surveillance. *Nucleic Acids Res.* **38**, 279–298 (2010).
- Wang, G. et al. PNPase regulates RNA import into mitochondria. *Cell* **142**, 456–467 (2010).
- Hornig-Do, H. T. et al. Isolation of functional pure mitochondria by superparamagnetic microbeads. *Anal. Biochem.* **389**, 1–5 (2009).
- Vedrenne, V. et al. Mutation in *PNPT1*, which encodes a polyribonucleotide nucleotidyltransferase, impairs RNA import into mitochondria and causes respiratory-chain deficiency. *Am. J. Hum. Genet.* **91**, 912–918 (2012).
- Matilainen, S. et al. Defective mitochondrial RNA processing due to *PNPT1* variants causes Leigh syndrome. *Hum. Mol. Genet.* **26**, 3352–3361 (2017).
- Rodero, M. P. et al. Detection of interferon alpha protein reveals differential levels and cellular sources in disease. *J. Exp. Med.* **214**, 1547–1555 (2017).
- Murr, C., Widner, B., Wirleitner, B. & Fuchs, D. Neopterin as a marker for immune system activation. *Curr. Drug Metab.* **3**, 175–187 (2002).
- White, M. J. et al. Apoptotic caspases suppress mtDNA-induced STING-mediated type I IFN production. *Cell* **159**, 1549–1562 (2014).
- Mannion, N. M. et al. The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).
- Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
- West, A. P. & Shadel, G. S. Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat. Rev. Immunol.* **17**, 363–375 (2017).
- Oda, H. et al. Aicardi-Goutières syndrome is caused by *IFIH1* mutations. *Am. J. Hum. Genet.* **95**, 121–125 (2014).
- Rice, G. I. et al. Gain-of-function mutations in *IFIH1* cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat. Genet.* **46**, 503–509 (2014).
- Crow, Y. J. & Manel, N. Aicardi-Goutières syndrome and the type I interferonopathies. *Nat. Rev. Immunol.* **15**, 429–440 (2015).
- Rutsch, F. et al. A specific *IFIH1* gain-of-function mutation causes Singleton-Merten syndrome. *Am. J. Hum. Genet.* **96**, 275–282 (2015).
- Lincez, P. J., Shanina, I. & Horwitz, M. S. Reduced expression of the MDA5 gene *IFIH1* prevents autoimmune diabetes. *Diabetes* **64**, 2184–2193 (2015).
- Gorman, J. A. et al. The A946T variant of the RNA sensor IFIH1 mediates an interferon program that limits viral infection but increases the risk for autoimmunity. *Nat. Immunol.* **18**, 744–752 (2017).

Acknowledgements We acknowledge V. Bondet for CSF IFN α data. We also thank E. Johnson and A. Pielach from the Dunn School Bioimaging facility for electron microscopy work and M. Alexeyev for sharing plasmids encoding UL12.5M185 and mUNG1 (Addgene #70109 and #70110, respectively). This work was supported by funding to N.J.P. (Wellcome Trust Investigator Award (107928|Z|15|Z), ERC Advanced Grant (339270)) and to M.T. (National Institutes of Health (NIH) (GM073981)). Y.J.C. acknowledges funding from the European Research Council (GA 309449: fellowship), a state subsidy managed by the National Research Agency (ANR, France) under the Investments for the Future (ANR-10-IAHU-01), and an ANR grant CE17001002 to Y.J.C. and D.D. Y.J.C. and D.D. thank ImmunoQure AG for sharing of the antibodies used to assess IFN α protein levels in the Simoa assay. Studies were supported by grant 2014/13/D/NZ2/01114 (to R.J.S.) from the National Science Centre, Poland. Experiments were carried out with the use of CePT infrastructure financed by the European Union through the European Regional Development Fund (Innovative economy 2007–13, Agreement POIG.02.02.00-14-024/08-00).

Reviewer information Nature thanks C. Reis e Sousa, S. Riis Paludan, G. Shadel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.Dh. conceived the study, designed and performed most of the experiments and drafted the paper. S.D. performed the bioinformatics analysis. L.S.B., R.J.S. and A.Dz. performed the experiments in Fig. 1c–f and Extended Data Figs. 1c, d, g, 2c–e, 3, 4. L.J. and M.T. generated the immunofluorescence and gene expression data on PNPase HepKO. A.R., A.M. and M.S. provided patient fibroblasts and the clinical data. C.T. provided the PNPase western blot (Fig. 3a). Y.J.C. and G.I.R. generated the blood ISG expression data. D.D. provided the IFN α CSF data. T.N. provided Extended Data Fig. 1e. C.R.A. performed FACS analysis. J.R. provided the MEF KO cells and agonists. A.Dh. discussed and interpreted results with inputs from N.J.P., R.J.S. and the other authors. N.J.P. and A.Dh. wrote the paper with input from the other authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0363-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0363-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.D. or R.J.S. or N.J.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

In vitro experiments were not randomized, in vivo experiments were randomized. The investigators were blinded to allocation during in vivo experiments (immunohistochemistry) and outcome assessment. No statistical methods were used to predetermine sample size.

Antibodies and reagents. The following antibodies were obtained commercially: mouse anti-ADAR1 mAb (sc-73408, Santa Cruz), rabbit anti-PNPT1 (ab96176, Abcam; sc-49315, Santa Cruz), mouse anti-dsRNA mAb J2 (10010500, Scions), anti-DNA (61014, Progen), rabbit anti-SUV3 (A303-055A, Bethyl Laboratories)¹¹, mouse anti-RIG-I mAb (Alme-1) (AG-20B-0009, AdipoGen), rabbit anti-COX IV (3E11, Cell Signaling), rabbit anti-cytochrome *c* (NB100-91732, Novus Biologicals), rabbit anti-calnexin (2433, Cell Signaling), mouse anti-lamin A/C (4C11, Cell Signaling), donkey anti-mouse IgG (H+L) conjugated with Alexa Fluor 488 (A-21202, Lifetech), goat anti-mouse IgM conjugated with Alexa Fluor 555 (A-21426, Thermo Fisher Scientific), normal mouse IgG2a (sc-3878, Santa Cruz), mouse anti-MDA5 (in house from J.R.), rabbit anti-MAVS (ALX-210-929-C100, Enzo Life Sciences), rabbit anti-Bax (2772T, Cell Signaling), rabbit anti-Bak (6947T, Cell Signaling), rabbit anti-OXA1L (HPA003531, Sigma), mouse anti- α -Tubulin (T5168, Sigma), mouse anti-actin (ab8226, Abcam; A5441, Sigma), rabbit anti-Flag (PA1-984B, Thermo Fisher Scientific), HRP secondary anti-mouse (ab6728, Abcam; A9044, Sigma)¹¹, HRP secondary anti-rabbit (A0545, Sigma; ab6721, Abcam)¹¹, goat anti-mouse IgG (20 nm gold) preadsorbed (ab27242, Abcam). ABT-737 (sc-207242, Santa Cruz), 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) (D1916, Sigma) or Act D (A1410, Sigma), digitonin (D141, sigma), MitoTracker Deep Red (M22426, Thermo Fisher Scientific), MitoTracker Red CMXRos (9082, Cell Signaling). ppp-IVT-RNA^{99nt} and CIP-EMCV-RNA were provided by J.R.²⁸

Cell culture, siRNA transfection and western blotting. HeLa (ATCC), hSUV3_{WT}/hSUV3_{G207V} HEK 293 cells¹¹, PNPase_{WT}/PNPase_{R445E-R446E} HeLa cells, or HEK 293 Flp-In T-Rex cells (Thermo Fisher Scientific), MEFs (*Ddx58*^{+/-}, *Ddx58*^{-/-}, *Ifih1*^{-/-})²⁹ and skin fibroblasts were grown as a monolayer at 37°C, under 5% CO₂ in DMEM (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fisher Scientific). Skin fibroblasts were isolated from skin biopsies of controls and PNPT1 patients. Fibroblast medium was supplemented with 2 mM L-glutamine, 2.5 mM pyruvate, 100 μ g ml⁻¹ streptomycin, 100 U ml⁻¹ penicillin at 37°C. Silencing of genes of interest was performed using stealthRNA or other siRNAs (Extended Data Table 2) with Lipofectamine RNAiMAX (Thermo Fisher Scientific) in HeLa cells according to the manufacturer's instructions. The stealthRNA oligonucleotides and siRNAs were used at a final concentration of 20 nM. For double siRNA treatments, each siRNA was used at final concentration of 20 nM. Cells were harvested three days after transfection unless stated otherwise. For Flp-In cells, expression of exogenous genes was induced by addition of tetracycline to the culture medium at a concentration of 25 ng ml⁻¹. For western blots, total protein cell extracts were prepared in lysis solution (10 mM Tris, 140 mM NaCl, 5 mM EDTA, 1% (v/v) Triton X-100, 1% (w/v) deoxycholate, 0.1% (v/v) SDS) except for fibroblasts where lysis solution (50 mM Tris, 300 mM NaCl, 10 mM MgCl₂, 0.5% NP-40, 2 mM DTT, Protease Inhibitor Cocktail 1X (Roche)) was used. Protein concentration was determined by the Bradford method. Protein extracts (30 μ g per lane) were separated by SDS-PAGE and transferred to a nitrocellulose membrane (Protran, Whatman GmbH). Western blotting was performed according to standard protocols.

Plasmid transfection and establishing of stable cell lines. Plasmid transfections were performed with TranIT2020 (Mirus) according to the manufacturer's instructions. HeLa cells were plated on glass coverslips 24 h after transfection with plasmids encoding UL12.5M185 and mUNG1, and after one day of culturing were subjected to an immunofluorescence procedure as described in the 'Immunofluorescence labelling' section. The stable inducible cell lines were established using plasmids pRS946 (PNPase_{WT}), pRS950 (PNPase_{R445E/R446E}) and HeLa Flp-In T-Rex cells (gift from M. Hentze) as described previously¹¹. The identity of HeLa Flp-In T-Rex cells was confirmed using STR profiling by DSMZ (Germany).

Mice. Hepatocyte-specific *Pnpt1*^{HepKO} (HepKO) mice were generated by breeding *Alb*^{CRE/WT}/*Pnpt1*^{neo-flox/neo-flox} with *Alb*^{WT/WT}/*Pnpt1*^{neo-flox/neo-flox} as described¹². Mice are housed, bred and studied in accordance with an approved protocol consistent with the UCLA Chancellor's Animal Research Committee (ARC) policies and procedures, as stated in Laboratory Animals in Teaching and Research (rev. 1998), the provisions of the NIH Guide for the Care and Use of Laboratory Animals and all applicable state and federal regulations.

Identification of PNPT1 mutations. Exome sequencing was performed on genomic DNA (1 μ g) isolated from blood leukocytes. Exons were captured by the in-solution enrichment methodology (SureSelect Human All Exon Kits Version 3, Agilent) using biotinylated oligonucleotide probe library (Human All Exon v3 50 Mb, Agilent). Each genomic DNA was then sequenced as paired-end 75 bases (Illumina HiSEQ2000, Illumina). After demultiplexing, sequences were aligned to

the reference human genome hg19 using the Burrows–Wheeler Aligner (v.0.7.12). Downstream processing was carried out with the Genome Analysis Toolkit (GATK 3.7), SAMtools (v.1.4), and Picard (v.2.9.0-1), following documented best practices (<https://www.broadinstitute.org/gatk/guide/topic?name=best-practices>). Variant calls were made with the GATK Unified Genotyper. The annotation process was based on the latest release of the Ensembl database (version 75), dbSNP (version 140), 1,000 genome project (version 2013/05/02), Gnomad (version 2.0.2) and EVS (version ESP6500SI-V2). Variants were annotated and analysed using the Polyweb software interface designed by the Bioinformatics platform of University Paris Descartes. Sequences were filtered against SNPs (>0.1% frequency) reported in public (dbSNP, 1,000 genomes and Exome Variant Server) and in-house databases including intergenic and non-coding region variants. Only homozygous variations were considered for patient 1, born to consanguineous parents, resulting in a list of 14 genes with only *PNPT1*, encoding a mitochondrial protein. Targeted exome sequencing using a panel of known genes for mitochondrial disorders was performed for patient 2 and two heterozygous *PNPT1* mutations were identified. Exome sequencing was performed for patient 3 and her non-consanguineous parents. The same filtering was used, identifying only one gene with two compound heterozygous mutations in *PNPT1*. DNA sequencing confirmed these mutations as well as their segregation with the disease in the families. All these variations were predicted to be deleterious by several software packages (Extended Data Table 1). Informed consent for diagnostic and research studies was obtained for all subjects in accordance with the Declaration of Helsinki protocols and all studies were approved by local Institutional Review Boards in Paris such as the human research participants ethics committee, Comité de Protection des Personnes, Ile de France II.

Targeted ISG RNA expression in total blood. Whole blood was collected into PAXgene tubes, total RNA extracted using a PreAnalytix RNA isolation kit and RNA concentration assessed using a spectrophotometer (FLUOstar Omega, Labtech). RT-qPCR analysis was performed using the TaqMan Universal PCR Master Mix (Applied Biosystems), and cDNA derived from 40 ng total RNA. To generate a standard six probe interferon score, TaqMan probes for *IFI27* (Hs01086370_m1), *IFI44L* (Hs00199115_m1), *IFIT1* (Hs00356631_g1), *ISG15* (Hs00192713_m1), *RSAD2* (Hs01057264_m1) and *SIGLEC1* (Hs00988063_m1) were used. The relative abundance of each target transcript was normalized to the expression level of *HPRT1* (Hs03929096_g1) and *18S* (Hs999999001_s1), and assessed with the Applied Biosystems StepOne Software v2.1 and DataAssist Software v3.01. For all six probes, individual data were expressed relative to a single calibrator. RQ (relative quantification) is equal to $2^{-\Delta\Delta C_t}$, that is, the normalized fold change relative to the control data. The median fold change of the six genes compared to the median of 29 previously collected healthy controls was used to create an interferon score for each individual, with an abnormal interferon score being defined as greater than 2 s.d. above the mean of the control group, that is, 2.466. The experiment was performed in triplicate from one blood sample obtained from each individual.

Quantification of IFN α in CSF by Simoa assay. Simoa IFN α assay was developed using a Quanterix Homebrew Simoa assay and two autoantibodies specific for IFN α isolated and cloned from two patients mutated in *APSI* (causing autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia, APECED) patients as recently described^{16,30}. The 8H1 antibody clone was used as a capture antibody after coating on paramagnetic beads (0.3 mg ml⁻¹), and the 12H5 was biotinylated (biotin/antibody ratio of 30:1) and used as the detector. Recombinant IFN α 17/ α 1 (PBL Assay Science) was used as a standard curve after cross-reactivity testing. The limits of detection were calculated by the mean value of all blank runs + 3 s.d. and was 0.23 fg ml⁻¹.

Reverse transcription and real-time qPCR analysis. Total RNA was treated with TurboDNase (Ambion) and reverse-transcribed using SuperScript Reverse Transcriptase III (Invitrogen) with oligo (dT)₂₀ for *INFB1* mRNA, *Ifih1* mRNA, L-mRNA (EMCV). qPCR was performed with 2 \times Sensimix SYBR mastermix (Bioligne) and analysed on a Corbett Research Rotor-Gene GG-3000 machine.

Immunofluorescence of HeLa cells infected with EMCV or transcription inhibitors and skin fibroblasts. HeLa cells were grown on a coverslip in a 6-well plate 24 h before treatment. HeLa cells were infected with EMCV at multiplicity of infection (MOI) of 1 for the indicated time point. For transcription inhibitor treatment, cells were treated with dimethylsulfoxide (DMSO) or Act-D or DRB at the indicated concentrations for 60 min. For J2 immunofluorescence on HeLa or skin fibroblasts, cells were incubated with MitoTracker Red CMXRos (100 nM) for 30 min at 37°C before fixing in 4% PFA in PBS. Cells were washed three times with PBS and permeabilized with 0.25% Triton X-100 in PBS. Cells were then washed with 0.05% Tween20–PBS and incubated with 3% BSA in PBS for 30 min at room temperature. Primary antibodies anti-dsRNA (J2) were used at 1:200 in 3% BSA in PBS for 1 h at room temperature. Cells were washed three times with 0.05% Tween20–PBS and then incubated with secondary donkey anti-mouse IgG (H+L) conjugated with Alexa Fluor 488 at (1:300) concentration. Cells were then

washed three times with 0.05% Tween20-PBS and twice with PBS, and mounted with Vectashield mounting media with DAPI (Vector Laboratories). Z-stack images were collected with a FluoView1000 confocal microscope (Olympus) using a UPLSAPO 60.0 \times / 1.35 oil objective. Images were analysed using ImageJ and prepared using OMERO software.

Immunofluorescence of nuclease treated samples. HeLa cells were plated on glass coverslips one day before fixation. Mitochondria-specific dye MitoTracker Deep Red (200 nM) was added to culture 1 h before fixation. Cells were washed twice with PBS and fixed in 5% (v/v) formaldehyde, 0.25% (w/v) Triton X-100 and Hoechst 33342 (2 μ g ml⁻¹) in PBS for 30 min in room temperature. Cells were washed three times with PBS. Following enzymes were used: RNase T1 (EN0541, Thermo Fisher Scientific, concentration 100 U ml⁻¹), RNase III (M0245S, NEB, concentration 40 U ml⁻¹), TURBO DNase (AM2238, Thermo Fisher Scientific, concentration 40 U ml⁻¹). Enzymes were added in PBS containing 5 mM MgCl₂. Samples were incubated in 37 °C for 30 min and washed three times with PBS. Cells were incubated with 3% (w/v) BSA in PBS for 30 min. Primary antibodies anti-dsRNA (2.5 μ g ml⁻¹) and anti-DNA (0.5 μ g ml⁻¹) were used in 3% (w/v) BSA 16 h at 4 °C. Cells were washed three times with PBS and secondary goat IgG anti-mouse IgG2a conjugated with Alexa Fluor 488 and goat anti-mouse IgM conjugated with Alexa Fluor 555 (Thermo Fisher Scientific) were used at 2 μ g ml⁻¹ concentration in 3% (w/v) BSA. Cells were incubated for 1 h at room temperature and washed three times with PBS and mounted. Slides were imaged with a FluoView1000 confocal microscope (Olympus) and with ScanR fluorescence microscopy system (Olympus) (using UPlanSApo 20.0 \times objective) adapted for high throughput image acquisition. The latter was used for quantitative fluorescent signal analysis. Quantification was performed for at least 1,755 cells per condition. Images were analysed using ScanR_2.7.2 analysis software (Olympus). The same microscope instrument settings were used for all samples.

RNA polymerase inhibition. For Extended Data Fig. 2e, HeLa cells were treated with siRNA for 3 days in 384-well format. Prior to fixation cells were treated for a given time with inhibitors of transcription: actinomycin D (0.5 μ g ml⁻¹), DRB (100 μ M). Detailed procedure is described in 'siRNA transfection in 384-well format' and 'Immunofluorescence labelling' sections. Quantitative analysis of dsRNA fluorescent signal was performed with ScanR fluorescence microscopy system. This analysis was performed for at least 500 cells per replica per condition. Images were analysed using ScanR_2.7.2 analysis software (Olympus).

siRNA transfection in 384-well format. Cells were reverse transfected in 384-well microplates (781946, Greiner Bio-One) using siRNA (final concentration 20 nM) and Lipofectamine RNAiMAX according to the manufacturer's instructions (Thermo Fisher Scientific). Cells were plated with the Multidrop Combi Reagent Dispenser (Thermo Fisher Scientific). After 72 h, cells were subjected to an immunofluorescence procedure described in the 'Immunofluorescence labelling' section. Cells were left in PBS for imaging. All PBS washes were performed with 405 LS Microplate Washer (BioTek) and all other solutions were added with the Multidrop Combi Reagent Dispenser.

Immunofluorescence labelling. One hour before fixation, mitochondria-specific dye MitoTracker Deep Red (200 nM) was added to the culture. Cells were washed twice with PBS and fixed for 30 min with PBS solution containing 5% (v/v) formaldehyde, 0.25% (w/v) Triton X-100 and Hoechst 33342 (2 μ g/ml). Cells were washed three times with PBS and incubated with 3% (w/v) BSA in PBS for 30 min. Primary antibodies against anti-dsRNA (2.5 μ g ml⁻¹) were used in 3% (w/v) BSA overnight at 4 °C. Cells were washed three times with PBS and secondary goat IgG anti-mouse IgG2a conjugated with Alexa Fluor 488 and goat anti-mouse IgM conjugated with Alexa Fluor 555 (Thermo Fisher Scientific) were used at 2 μ g ml⁻¹ concentration in 3% (w/v) BSA. Cells were incubated 1 h at room temperature and washed three times with PBS. Cover slips were mounted with ProLong Gold Antifade Mountant (P36930, Thermo Fisher Scientific) or left in PBS if imaged with a ScanR fluorescence microscopy system. If the samples were subjected to quantitative analysis, the same microscope instrument settings were applied.

Co-localization of dsRNA with mitochondria. Cells were subjected to staining as described in the 'Immunofluorescence labelling' section. Z-stack images of microscopic slides were collected with a FluoView1000 confocal microscope (Olympus) using a PLANAPO 60.0 \times / 1.40 oil objective. XY optical resolution of images was 215 nm. Images were analysed using Imaris v7.2.3 software (Bitplane). Object-based colocalization of spots was performed. Colocalization of J2 spots with mitochondria was based on fluorescence intensity from MitoTracker. Quantification was performed for 29 randomly selected cells.

High-throughput fluorescence imaging. Data presented on Fig. 1f, Extended Data Figs. 2c, d, 4d were obtained using a ScanR fluorescence microscopy system (Olympus, UPlanSApo 20.0 \times objective). Images were analysed using ScanR 2.7.2 analysis software (Olympus). Quantification of fluorescent signal was performed for at least 400 cells per replica per condition.

Fluorescent immunohistochemistry. Fluorescent immunohistochemistry staining of 4 μ m-thick formalin-fixed, paraffin-embedded (FFPE) tissue sections was

performed on livers from sex-matched (female) six-week-old wild-type C57BL/6 and *Pnpt1*^{HepKO} (HepKO) littermate mice on a pure C57BL/6 background. FFPE slides were deparaffinized by immersion in 100% xylene, two times for 5 min and rehydrated twice in fresh 100% ethanol, 95% ethanol, 70% ethanol, and 50% ethanol for 3 min each. Sections were washed in double distilled H₂O and permeabilized with 0.1% Triton X-100 in 1 \times PBS. Heat-induced antigen retrieval (HIER) was performed by heating sections in 1 mM EDTA at 95 °C in a pressure cooker for 20 min and then 20 min of cooling at room temperature. Sections were incubated for 12 h at 4 °C in blocking buffer (5% goat serum + 0.3% Triton X-100 + 3% BSA diluted in 1 \times PBS) and subsequently incubated with mouse anti-dsRNA J2 antibody, diluted 1:200 in blocking buffer, for 2 h at room temperature. Sections were washed in 0.1% Triton X-100 in 1 \times PBS three times for 10 min and incubated with secondary antibody goat anti-mouse Alexa Fluor 488 at 1:200 for 1 h at room temperature and washed again with 1 \times PBS-Tween. All processed slides were mounted in Prolong gold antifade mount with DAPI (Invitrogen cat # P36931). Some slides were processed in the absence of primary antibody to verify specificity of labelling.

Imaging of immunohistochemistry samples. All images were obtained with a 100 \times objective on a Leica TCS-SP8 inverted spectral confocal microscope (Leica Microsystems) equipped with a 405 nm blue diode laser, argon laser (5 lines), and white light laser for excitation. Further image processing of maximal z-projection images of 4 μ m thick liver sections showing 488 and DAPI overlay was performed on LAS X v.3.30 software. Identical settings were used to obtain fluorescent images within datasets. Brightness and contrast for final images were adjusted equally across datasets using Photoshop CC 2017. Confocal laser-scanning microscopy was performed at the CNSI Advanced Light Microscopy/Spectroscopy Shared Resource Facility at UCLA.

Quantification of immunohistochemistry samples. To quantify J2 dsRNA immunofluorescence in liver sections from wild-type C57BL/6 and *Pnpt1*^{HepKO} (HepKO) mice, a single in-focus plane was acquired at 100 \times at 20–21 locations across the tissue selected using a random coordinate generator. Quantifications were performed using ImageJ software, by drawing an outline around tissue and measuring area, integrated density and mean fluorescence. Additionally, background readings were measured on secondary-only tissue samples. To calculate the corrected total fluorescence intensity (CTFI) we used the following formula: CTFI = integrated density – (area of selected tissue \times mean fluorescence of background readings). Scatter plot and statistical analysis (two-sided unpaired *t*-tests with Welch's correction) were performed using GraphPad Prism 7.

Flow cytometry analysis. Cells were trypsinized, washed with PBS and fixed with 4% formaldehyde diluted in PBS for 20 min at room temperature. Cells were permeabilized with 0.1% Triton X-100 in PBS for 15 min followed by incubation in 1% BSA (Sigma, A7030) in PBS for 1 h. Primary antibodies (anti-dsRNA (J2) or normal mouse IgG2a (Iso)) were used at 2.5 μ g ml⁻¹ in 1% BSA for 1 h at room temperature. Secondary donkey anti mouse IgG (H+L) conjugated with Alexa Fluor 488 were used at 2.2 μ g ml⁻¹ concentration for 1 h at room temperature. Cells were rinsed three times with FACS buffer (0.5% BSA in PBS with 2 mM EDTA). Data were acquired with a FACSCalibur (BD Biosciences) flow cytometer. Data were analysed in FlowJo (TreeStar).

Immunoprecipitation of dsRNA. Protein G Dynabeads were washed and resuspended in NET-2 buffer. 5 μ g of anti-dsRNA mAb (J2) was bound to 100 μ l of beads for 1 h at room temperature on a Thermoshaker. Conjugated beads were washed three times with NET-2 Buffer. 80–90% confluent HeLa cells from 10 cm² plate (\times 2) were washed with 10 ml of cold PBS. Cells were scraped and transferred to a falcon and spun at 500g at 4 °C, 5 min. Cell pellet from one 10-cm² plate was lysed in 1 ml of NP-40 lysis buffer and transferred to an eppendorf and incubated on ice for 5 min. Following a spin at 17,000g at 4 °C for 5 min, supernatant was carefully transferred to a new eppendorf. Total RNA was harvested from 10% input lysate using Trizol reagent. For immunoprecipitation, lysate was diluted 1:4 in NET-2 buffer and supplemented with 10 units of RNase free TurboDNase (Ambion) at 10 mM MgCl₂ per 1 ml of mix. 100 μ l of J2-Dynabeads was added to 1 ml of above lysate and left for 1–2 h at 4 °C. Following magnetic separation, beads were washed twice with 1 ml of high salt washing buffer (HSWB). Beads were transferred to a new tube with NET-2 buffer and washed twice with the same buffer. J2-bound dsRNA was extracted with Trizol reagent. The RNA samples were sent for sequencing (described in the 'RNA-sequencing' section). NET-2 buffer (50 mM Tris-Cl, pH 7.4, 150 mM NaCl, 1 mM MgCl₂, 0.5% NP-40), NP-40 lysis buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40), high salt wash buffer (50 mM Tris-Cl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.5% DOC, 0.1% SDS). **dsRNA isolation for northern blot.** HEK 293 cells from 150-cm² plate (\times 3) were used. Cell pellet was lysed in 4.5 ml of NP-40 lysis buffer and kept on ice for 5 min. Lysate was transferred to 1.5 ml eppendorf and centrifuged at 20,000g at 4 °C for 5 min. The supernatant was then transferred to 15 ml tube. Lysate was diluted 1:4 in NET-2 buffer and supplemented with 12 units of RNase free TurboDNase (Ambion) and 10 mM MgCl₂ per 1 ml of final mix. RNases were added (RNase

T1-1U from 1 U μl^{-1} , RNase V1-1U from 0.1 U μl^{-1} (Life Technologies)) and incubated at 37°C for 10 min. 100 μl of J2-Dynabeads were added to 15 ml of above lysate and left on a rotor at 4°C for 1–2 h. Beads were spun at 3,000g at 4°C, 3 min. Supernatant was discarded and beads transferred to 1.5 ml tube, washed twice with 1 ml of HSWB and washed twice with NET-2 buffer. J2-bound dsRNA was extracted with Trizol reagent. RNA samples were used for northern blot.

Northern blot analysis. dsRNA after immunoprecipitation was purified by TRI Reagent (Sigma) using the manufacturer's protocol. 20% of dsRNA eluate was dissolved in denaturing solution and run on a 1% agarose gel as described previously¹¹. Subsequently, RNA was transferred to Amersham Hybond-N+ membrane (GE Healthcare Life Sciences) and UV cross-linked. For detection of mitochondrial transcripts probes were labelled with [α -³²P] dATP (Hartmann Analytic) using a DECAprime II Kit (Ambion). PCR products corresponding to the following fragments of human mtDNA were used as templates: 254–4469 (Probe 1), 4470–8365 (Probe 2), 8365–12137 (Probe 3), 12091–16024 (Probe 4). Hybridizations were performed in PerfectHyb Plus buffer (Sigma) at 65°C. Membranes were exposed to PhosphorImager screens (FujiFilm), which were scanned following exposure by a Typhoon FLA 9000 scanner (GE Healthcare). Data were analysed by Multi Gauge v3.0 software (FujiFilm).

Probes for RNA protection assay (RPA). U1 snRNA antisense fragment was amplified from pGEM4-tU1 (S. Murphy, University of Oxford) by PCR using the following primers: tU1 forward, AGCTCGGATCCATACCTACCTGGCAGGGGAGATA; tU1 reverse, ATTCATTAATGCAGCTGGCTT. According to the manual of StrataClone Blunt PCR cloning kit (Agilent Genomics), the PCR product was cloned as pSC-B-tU1_RPA. T7 transcription was performed using [α -³²P]UTP and XhoI-digested pSC-B-tU1_RPA to label the antisense tU1 RNA. The radio-labelled RNA was purified from 6% denaturing gel.

In vitro J2 immunoprecipitation assay. In brief, 2,000 cps [α -³²P]UTP-labelled antisense tU1 RNA was incubated with 10 μg of purified HeLa nuclear RNA followed by RNase protection analysis (RPA)³¹. After RPA, dsRNA was immunoprecipitated with 5 μg of J2 antibody conjugated protein G beads (Thermo Fisher Scientific) in NET-2 buffer. [α -³²P]-labelled antisense tU1 RNA was used as ssRNA substrate. The antibody beads were washed with NET-2 buffer several times and then incubated with Trizol (Thermo Fisher Scientific) to purify immunoprecipitated RNAs. The RNAs were analysed on 8% denaturing gel.

mtRNA isolation and treatment of cells. Mitochondria were isolated from HeLa cells using magnetic cell separation procedure as described by the manufacturer (Mitochondria Isolation Kit; MACS, Miltenyl Biotec). RNA was purified from mitochondria using Trizol reagent (Sigma) and was treated with RNase-free TurboDNase (Ambion) according to manufacturer's instructions. 1 μg of mtRNA was transfected into HeLa cells and 300 ng was transfected into MEFs in a 1:3 ratio with Lipofectamine 2000 in 12-well plates with cells at 80% confluency. For enzymatic treatment, 1 μg of mtRNA was incubated with RNase III as per the manufacturer's instructions. 100 ng of ppp-IVT-RNA^{99nt} and CIP-EMCV-RNA were transfected in MEFs in 12-well plates using Lipofectamine 2000 in a 1:3 ratio. Total RNA from HeLa cells or MEFs was extracted 20 h after transfection for *IFNB1* or *Ifit1* mRNA quantification, respectively. In Extended Data Fig. 6b, HeLa cells were treated with ABT-737 (10 μM) or DMSO 65 h after siRNA transfection and incubated for a further 8 h. Total RNA was isolated using Trizol for *IFNB1* mRNA quantification.

Separation of cytoplasm and mitochondria fractions. Mitochondrial and cytoplasmic fractions in Extended Data Fig. 6c were prepared as previously described³². Purity of fractions was tested by western blot according to standard protocols.

Detection of mtRNA in cytosolic extracts. Cytosolic extracts (Fig. 3c) were prepared using digitonin extraction as described previously⁸. Digitonin (Sigma) at 25 $\mu\text{g ml}^{-1}$ was used for control and patient fibroblast cells. Purity of cytosolic fractions was tested by western blot, Lamin A/C was probed as a nuclear loading control, Calnexin was probed for endoplasmic reticulum, COX IV was probed for mitochondria and Tubulin was probed as a cytoplasmic control. RT-qPCR was performed on RNA isolated from cytosolic fractions using random hexamers for cDNA synthesis followed by qPCR using mtDNA-specific primers (Extended Data Table 2) normalized to *ACTB* mRNA levels.

RNA sequencing. RNA sequencing was performed by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, University of Oxford. Input RNA samples were ribo-depleted with Ribo-Zero rRNA-removal kit (Human/Mouse/Rat, EpiCentre RZH110424). Immunoprecipitated RNA samples were not ribo-depleted. Libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina, v1.0 (cat. no. E7420) according to the manufacturer's guidelines. Libraries were sequenced on an Illumina HiSeq-2000 with 100-bp paired-end reads, v3 chemistry.

Microarray method. Hepatocytes were isolated from perfused livers of two PNPase (*Pnpt1*) liver-specific knockout C57BL/6 mice (HepKO; one male aged 12.9 weeks, one female aged 4.29 weeks, two independent experiments) and two sex-matched wild-type littermate mice. Total RNA was extracted from hepatocytes

using TriZol Reagent (Invitrogen cat. #15596026) and the Qiagen RNeasy Mini Kit (Qiagen, cat. #74104). Labelled complementary RNA was generated using 200 ng of total RNA from each sample and the Agilent Low RNA Input Linear Amplification labelling kit. Each labelled sample was hybridized against its gender-matched sample in fluor-reversed pairs of arrays to an Agilent 4×44k Mouse Whole Genome Microarray. The arrays were scanned using the Agilent DNA Microarray Scanner, and data were extracted using the Agilent Feature Extraction (v9.5.1.1) software using the standard Agilent protocol except without Lowess normalization. The fluor-reversed pairs were combined into Experiments in Rosetta Resolver 7.1 to produce the male and female signature gene ratios. We performed age and gender-matched differential expression analysis and generated a list of signature genes with significant differential expression in both the male and female cohorts. Both male and female cohorts showed very similar results. For simplicity, only fold expression changes in female mice were shown.

Interferon reactome methods. The gene set for interferon signalling (encompassing IFN $\alpha/\beta/\gamma$ signalling and interferon-stimulated antiviral response) was extracted from the Reactome database under pathway ID R-MMU-913531³³. An additional antiviral innate immune response gene set was curated from recent studies investigating the role of mitochondrial DNA in innate immunity⁸. Corresponding genes from the set were compared to the HepKO signature genes (adjusted $P < 0.05$, significant in both female and male matched pairs). Overlapping genes were plotted using the male and female matched log₂(fold change). Both male and female data showed very similar results. For simplicity, only fold expression change data from female mice were shown.

Immuno-electron microscopy. HeLa cells were grown in a 6-well plate and siRNA treated for 65 h, trypsinized and pelleted for 1 min at 1,000g in cell culture media containing 5% FBS and 20% BSA. Cells were then immediately cryo-fixed using a Leica EM PACT2 high pressure freezer and then further processed as described³⁴, except that tannic acid was omitted from the freeze substitution medium. Blocks were sectioned on a Leica UC7 ultramicrotome using a diamond knife (Diamtome). 90-nm sections were transferred to 200-mesh Nickel grids and then immunolabelled as follows: grids were floated section side down on a 20 μl droplet of blocking solution (10% goat serum in TBS) for 15 min, then blotted and incubated on a droplet of primary (J2) antibody (diluted 1:25 in buffer A (1% BSA, 1% goat serum, 0.01% Tween-20 in TBS)) for 2 h at room temperature. Grids were washed by passing them over five droplets of buffer A, 5 min each, then incubated with secondary antibody (ab27242, goat anti-mouse conjugated to 20 nm gold) diluted 1:10 in buffer A for 90 min at room temperature, then washed by passing over three droplets of water. Sections were then post-stained for 10 min with uranyl acetate and Reynold's lead citrate and imaged on a Tecnai 12 transmission electron microscope (FEI) equipped with a Gatan OneView CMOS camera.

Statistical analysis. Unless otherwise stated, the figures present the mean values of at least three independent experiments with s.d. or s.e.m. For analyses with $n > 10$, individual data points are shown. The mean is reported when $n = 2$, and no other statistics were calculated for these experiments.

Bioinformatics: mapping of sequencing reads and data visualization. Paired-end reads for each sample were mapped to the human genome reference assembly GRh37/hg19 (build 37.2, Feb 2009) using the Bowtie2 alignment software³⁵. Prior to alignment, adaptor sequences were trimmed using Cutadapt 1.8.3, discarding reads with less than ten bases. An in-house Perl script was used to remove the reads left unpaired (code available upon request). SAMtools 1.2³⁶ was used to process aligned reads to only include uniquely mapped reads with no more than two mismatches. Data were scaled to library size (genomeCoverageBed) using Bedtools³⁷. Bigwig track files were generated from the Bowtie2 output files using UCSC bedGraphToBigWig tool³⁸. Correlations among different samples for chrM were calculated with R. Data from replicates ($n = 2$ for each condition) except untreated and control-siRNA-treated samples ($r = 0.85$) were then merged and viewed on the UCSC genome browser (Extended Data Fig. 5a, b). For the chromosome-wide read coverage plot of dsRNA-seq, the number of filtered reads mapping for each chromosome were counted. These numbers were then normalized to the size of the respective chromosome. For plotting the distribution of reads belonging to different RNA species, reads mapped to each Ensembl biotype annotations were counted for using bedtools and then normalized to the size of the genome. The heat map was generated with the R package gplots using a subset of significantly altered ISGs identified by gene expression analysis on total RNA sequencing (input RNA) from control siRNA (siRNA targeting the luciferase gene) and two knockdown conditions (siSUV3 and siPNPase).

To screen for RNA editing candidates in chrM, the REDIttoolDenovo.py script from the REDIttools package³⁹ was used. To avoid the risk of using unreliable editing sites, the output of REDIttools was passed through two stringent filters. First, all editing sites were removed that had less than ten edited counts and a total read coverage < 50 for all the knockdown libraries. Second, only those sites in which the frequency of editing by siRNA targeting SUV3 and PNPase was at least 1.5-fold that of control siRNA were considered. The resulting editing sites were

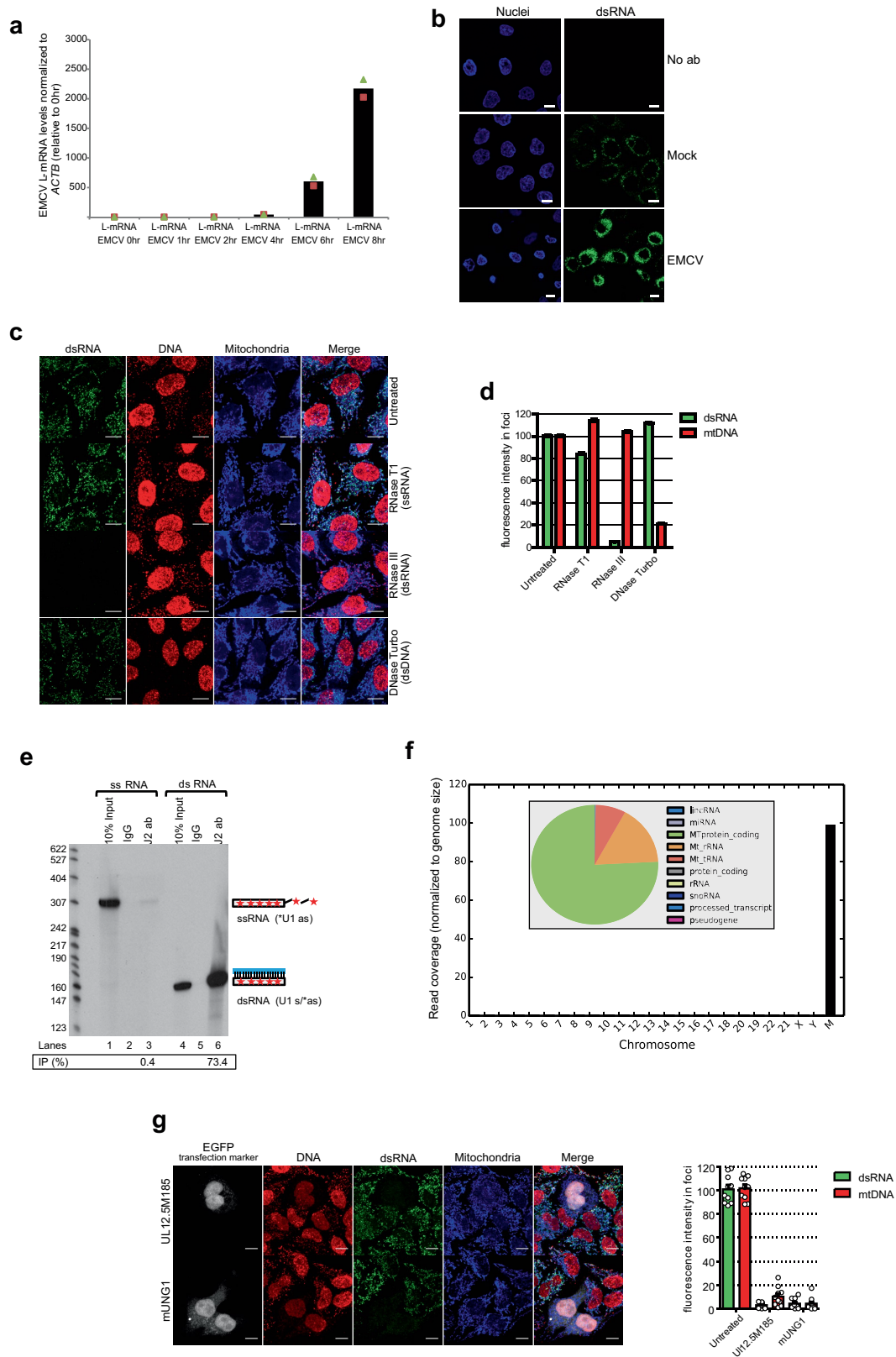
then filtered for known SNPs in the mitochondrial transcriptome⁴⁰ to obtain reliable RNA editing candidates.

Quantification of Neopterin levels in CSF. CSF pterins were analysed by reverse-phase high-performance liquid chromatography with fluorescence detection according to⁴¹.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Mouse microarray data have been deposited in the Gene Expression Omnibus (GEO) under accession numbers GSE94957 and GSE109210. Source Data for graphical representations obtained from the PNPase HepKO mouse microarray (Fig. 2d, e, Extended Data Fig. 6d) are available with the online version of the paper. Gel source images are presented in Supplementary Fig. 1.

28. Hertzog, J. et al. Infection with a Brazilian isolate of Zika virus generates RIG-I stimulatory RNA and the viral NS5 protein blocks type I IFN induction and signaling. *Eur. J. Immunol.* <https://doi.org/10.1002/eji.201847483> (2018).
29. Deddouche, S. et al. Identification of an LGP2-associated MDA5 agonist in picornavirus-infected cells. *eLife* **3**, e01535 (2014).
30. Meyer, S. et al. AIRE-deficient patients harbor unique high-affinity disease-ameliorating autoantibodies. *Cell* **166**, 582–595 (2016).
31. Nojima, T., Dienstbier, M., Murphy, S., Proudfoot, N. J. & Dye, M. J. Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell Reports* **3**, 1080–1092 (2013).
32. Nishimura, N. & Yano, M. Separation of the inner and outer mitochondrial membrane in HeLa cells. *Bio Protoc.* **4**, e1299 (2014).
33. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
34. Johnson, E. et al. Correlative in-resin super-resolution and electron microscopy using standard fluorescent proteins. *Sci. Rep.* **5**, 9583 (2015).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
36. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
38. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
39. Picardi, E. & Pesole, G. REDIttools: high-throughput RNA editing detection made easy. *Bioinformatics* **29**, 1813–1814 (2013).
40. Hodgkinson, A. et al. High-resolution genomic analysis of human mitochondrial RNA sequence variation. *Science* **344**, 413–415 (2014).
41. Blau, N. et al. Cerebrospinal fluid pterins and folates in Aicardi-Goutieres syndrome: a new phenotype. *Neurology* **61**, 642–647 (2003).

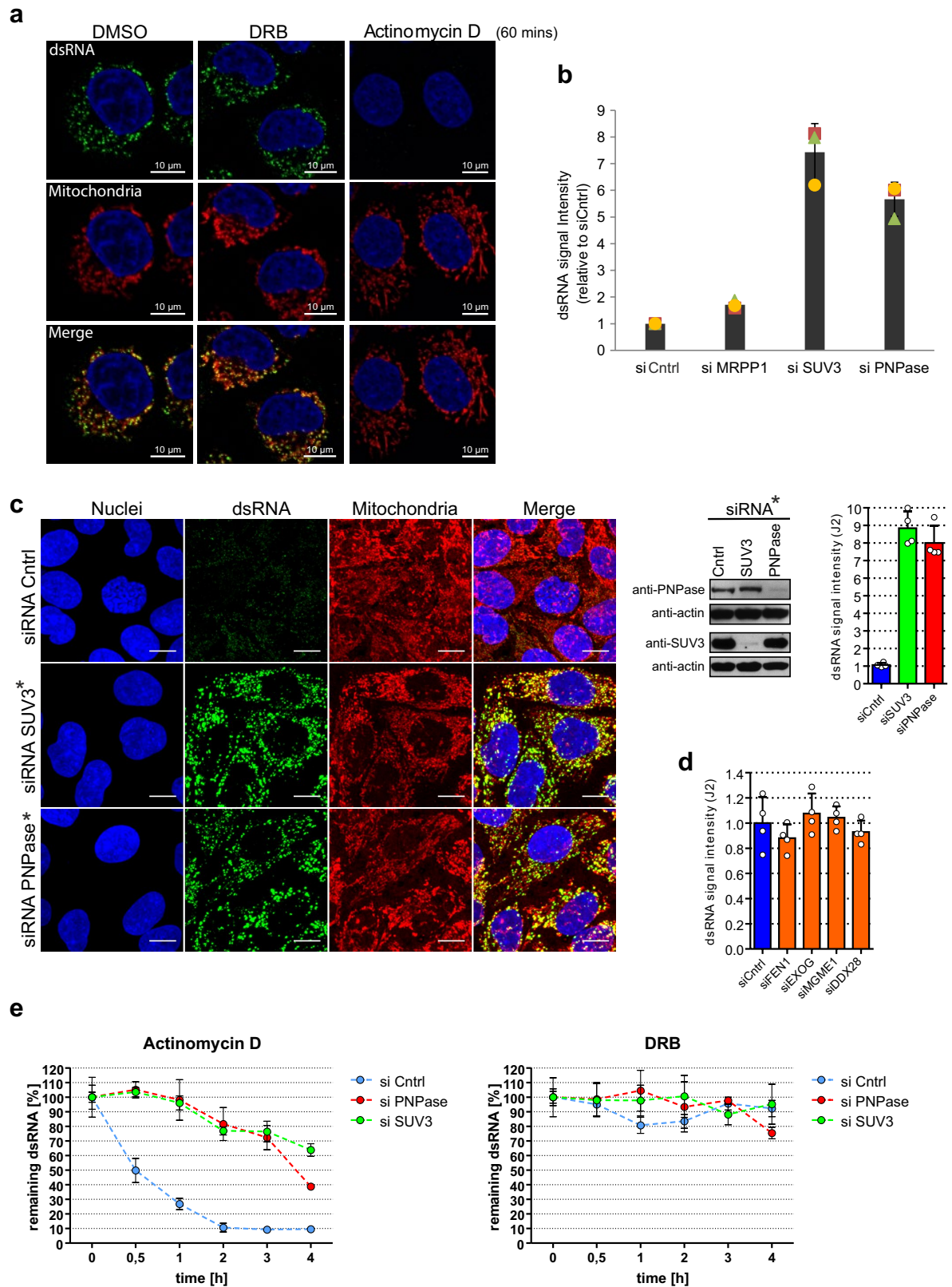


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Characterization of anti-dsRNA J2 antibody and mtDNA depletion results in loss of mtdsRNA formation.

a, RT-qPCR analysis of L-mRNA expression in encephalomyocarditis virus (EMCV) infected HeLa cells at MOI 1 at the indicated time points after infection. Data are from two independent experiments. **b**, Confocal microscopy images of uninfected or EMCV-infected HeLa cells at multiplicity of infection (MOI) of 1, 8 h after infection stained with anti-dsRNA (J2) antibody (green) and DAPI (nuclei stained blue). Images are representative of two experiments. Scale bars, 10 μm . **c**, Immunostaining of dsRNA (green) and DNA (red) in HeLa cells treated with indicated nucleases before staining. Signal from J2 antibody is specific for RNA but not for DNA and is sensitive only to RNase III treatment. Images are representative of three experiments. Scale bars, 10 μm . **d**, Quantification of fluorescence signal from HeLa cells treated as in **c**. Data are mean \pm s.e.m. from 4,095, 1,755, 4,766 and 5,585 cells for the untreated, RNase T1,

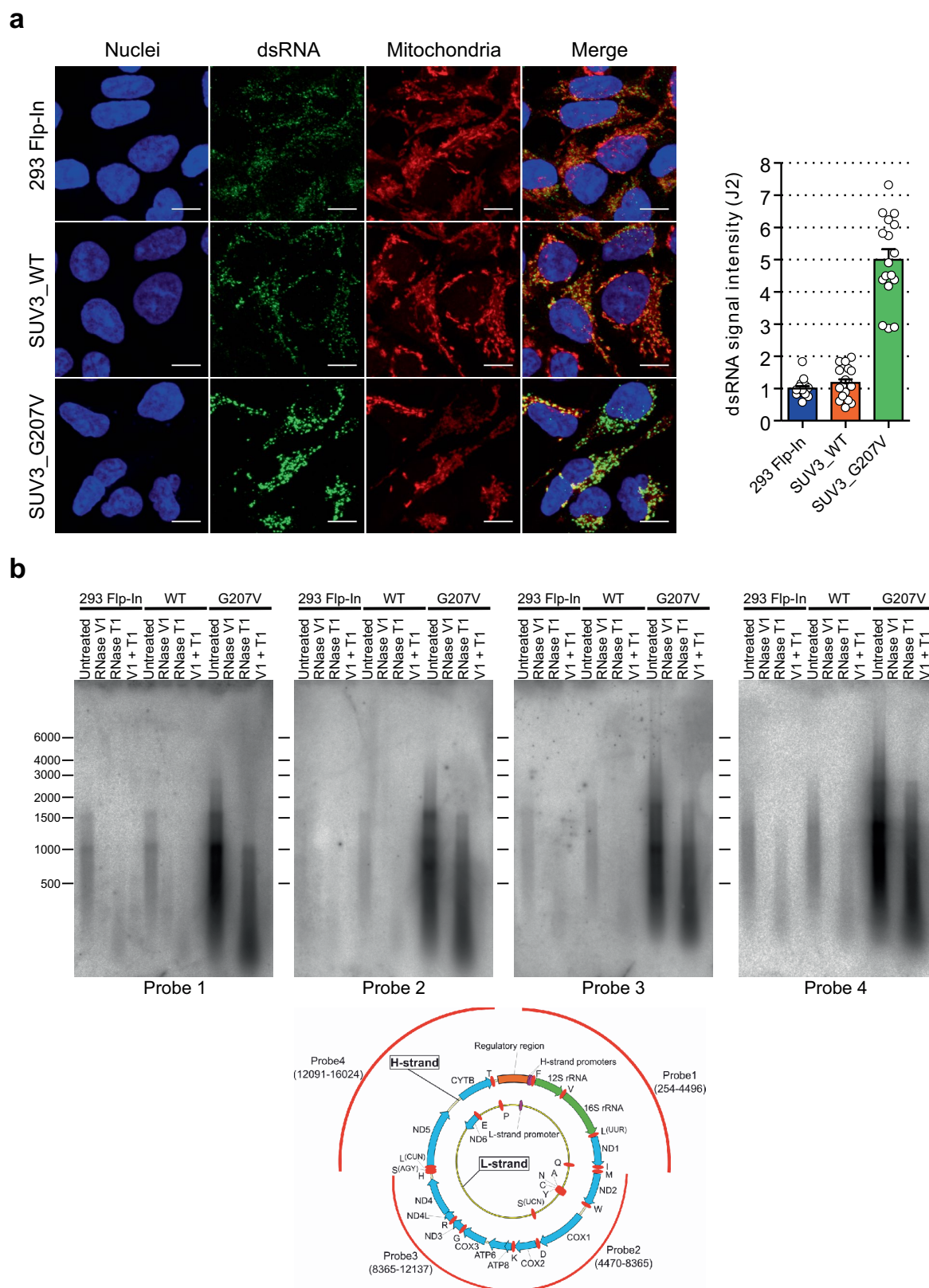
RNase III and DNase Turbo groups, respectively. **e**, Autoradiogram showing substrate specificity of J2 on the basis of immunoprecipitation efficiency for uniformly ^{32}P -radiolabelled ssRNA and dsRNA substrates. Signals were visualized and quantified by PhosphorImager. The level of immunoprecipitation signal is shown and expressed as the percentage of input. Images and data are representative of two experiments. For gel source data, see Supplementary Fig. 1. **f**, Chromosome-wise coverage plot of dsRNA-seq reads. Inset, read distribution of dsRNA-seq on the basis of RNA class biotypes. **g**, Left, dsRNA and DNA staining of HeLa cells transfected with constructs encoding the indicated proteins, the expression of which results in mtDNA depletion. Plasmids encoding mtDNA-depletion factors co-express EGFP from an independent promoter, which enables identification of transfected cells. Mitochondria were stained using anti-OXA1L antibody. Scale bars, 10 μm . Right, quantitative analysis of fluorescence signal from HeLa cells. Data are mean \pm s.e.m. from ten cells.



Extended Data Fig. 2 | See next page for caption.

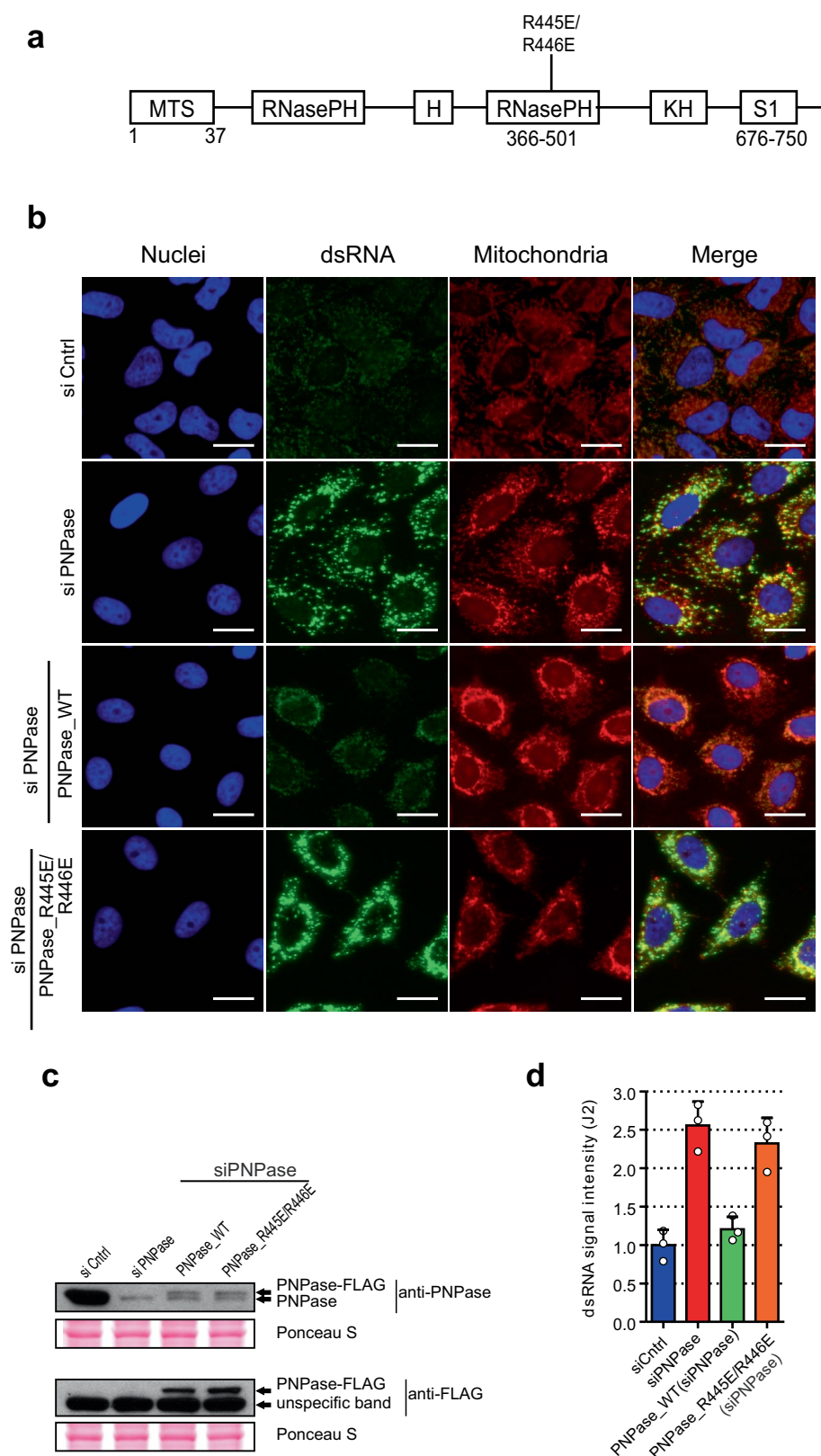
Extended Data Fig. 2 | RNA degradosome components SUV3 and PNPase are involved in mtdsRNA turnover. **a**, HeLa cells treated with DMSO, DRB (100 μ M) and actinomycin-D (0.5 μ g ml⁻¹) for 60 min and stained with anti-dsRNA (J2) antibody (green). Mitochondria were stained with MitoTracker Red CMXRos and nuclei with DAPI (blue) (representative of two experiments). **b**, Flow cytometric analysis of dsRNA levels in HeLa cells treated with the indicated siRNAs. Cells were labelled with J2 antibody or an isotype control. Data are mean \pm s.d. from three independent experiments. **c**, Left, detection of dsRNA with J2 antibody in HeLa cells after depletion of PNPase or SUV3 by On-TARGETplus siRNAs (indicated with an asterisk and listed in Extended Data Table 2). Mitochondria were stained with MitoTracker Deep Red. Nuclei are stained with Hoechst (blue). Scale bars, 10 μ m. Right top, western blot

showing PNPase or SUV3 depletion. Blots are representative of four experiments. For gel source images, see Supplementary Fig. 1. Far right top, Quantification of dsRNA levels in HeLa cells depleted of PNPase or SUV3. Data are mean \pm s.d. from four independent experiments. **d**, Quantitative analysis of fluorescent signals from dsRNA in HeLa cells with depleted enzymes involved in mitochondrial nucleic acids metabolism. Data are mean \pm s.d. from four independent experiments. **e**, HeLa cells were transfected with siRNA specific for PNPase, SUV3, or non-targeting control. Prior to fixation, cells were treated for indicated times with inhibitors of transcription: actinomycin-D (0.5 μ g ml⁻¹) or DRB (100 μ M). Immunostaining of dsRNA was performed and cells were imaged using a fluorescent microscope screening station. Data are mean \pm s.d. from four independent experiments.



Extended Data Fig. 3 | Unwinding activity of SUV3 is required to suppress mtdsRNA accumulation. **a**, Left, confocal images of HEK 293 cells expressing stably integrated wild-type SUV3 (hSUV3₃-WT) and the catalytically inactive (G207V) dominant negative form (hSUV3₃-G207V) stained with J2 ab (green). Mitochondria stained with MitoTracker Deep Red (red). Nuclei stained with Hoechst (blue). Right, quantitative analysis of fluorescence signal from HEK 293 cells in the above experiment.

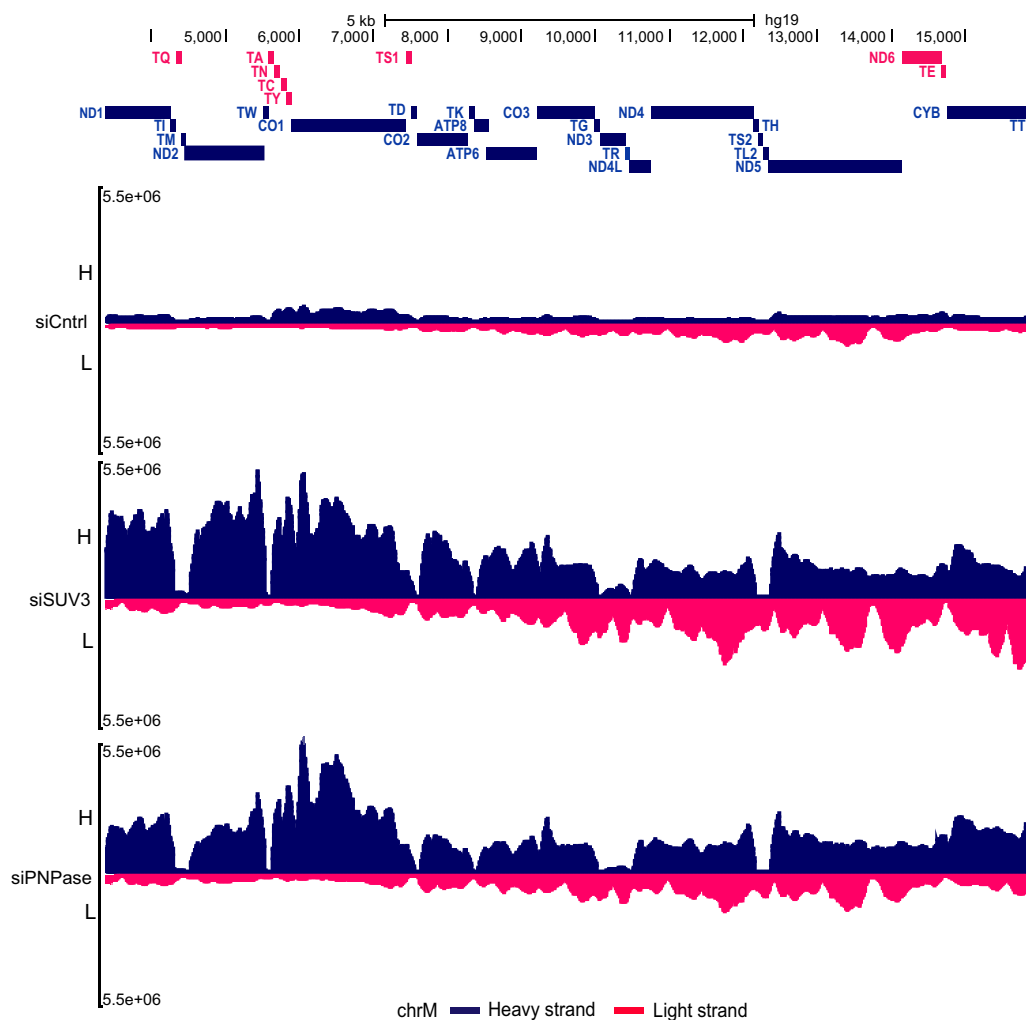
Data are mean \pm s.e.m. from 16 cells. **b**, Top, northern blots of J2 immunoprecipitated dsRNA from hSUV3_WT and hSUV3_G207V overexpressing HEK 293 cell lines with four different probes spanning the entire mitochondrial genome. Bottom, diagram depicts positions of probes on mitochondrial genome. Blots are representative of two experiments. For gel source data, see Supplementary Fig. 1.



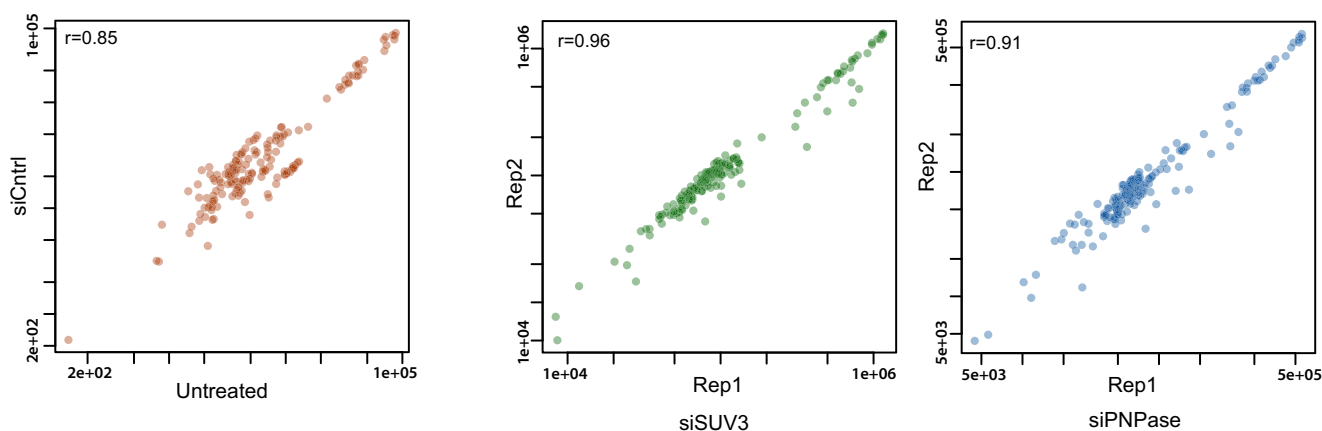
Extended Data Fig. 4 | Exonuclease activity of PNPase is required to suppress mtdsRNA formation. **a**, Diagram of PNPase domain structure showing the position of the R445E/R446E mutation in the RNasePH domain. **b**, Immunostaining of dsRNA (green) in HeLa stable cell lines transfected with siRNA specific for PNPase or non-targeting control siRNA. Depletion of endogenous PNPase was rescued by expression of siRNA-resistant PNPase-FLAG protein (wild-type or mutated (RNA-degradation deficient version of PNPase (R445E/R446E) was expressed)).

Mitochondria are stained with MitoTracker Deep Red. Nuclei are stained with Hoechst (blue). Scale bars, 20 μ m. **c**, Western-blot analysis of PNPase in HeLa cells treated as in **b**. Exogenous, siRNA-resistant PNPase is expressed as a FLAG fusion. Blots are representative of three experiments. **d**, Quantitative analysis of fluorescent signals from dsRNA in HeLa treated as in **b**. Data are mean \pm s.d. from three independent experiments. For gel source data, see Supplementary Fig. 1.

a

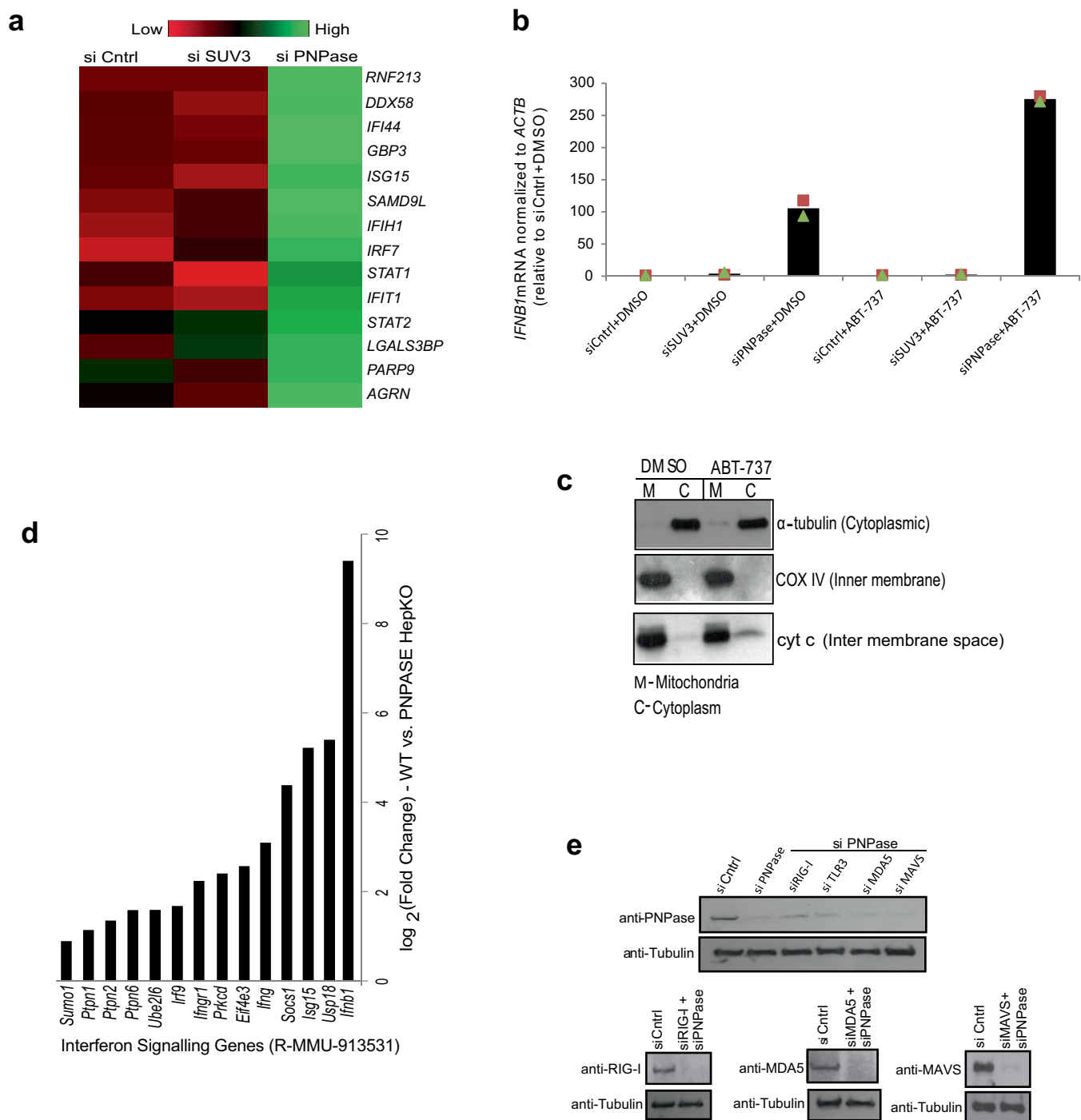


b



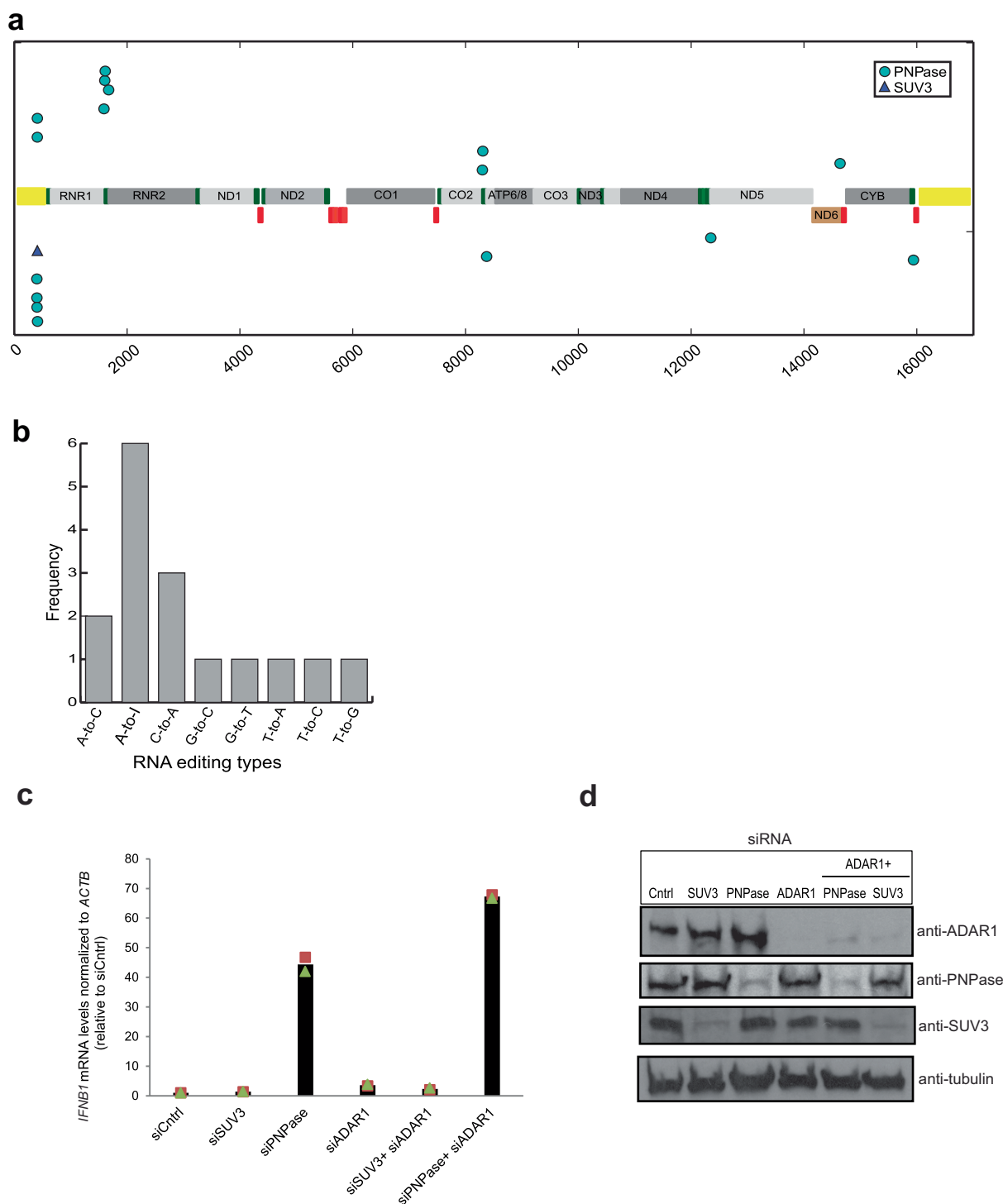
Extended Data Fig. 5 | dsRNA-seq of HeLa cells following siRNA depletion of SUV3 and PNPase. a, dsRNA-seq reads across the mitochondrial genome spanning entire protein coding region (~3.5–16 kb) following siRNA treatment. Data are from two independent experiments.

H-strand genes are shown as blue bars and L-strand as red bars. Short tRNA genes are denoted with T as the first letter. **b**, Correlation plots of J2 immunoprecipitation dsRNA-seq replicates. Pearson correlation coefficients are calculated and shown on each plot.



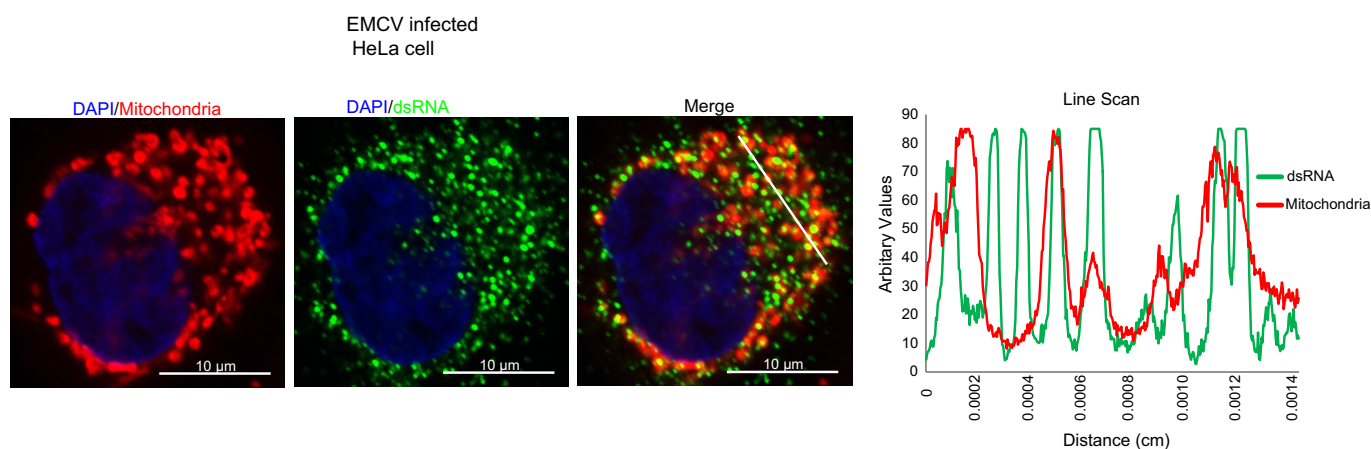
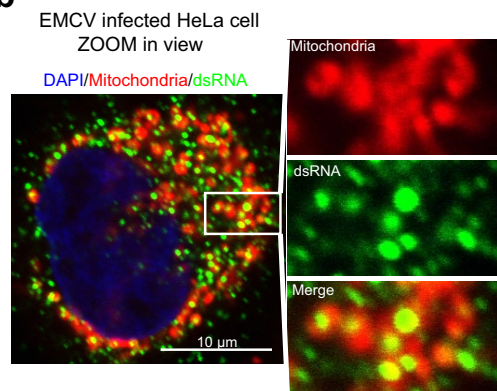
Extended Data Fig. 6 | Upregulation of ISGs in HeLa and murine cells following loss of PNPase accentuated by mitochondrial outer membrane permeabilization. **a**, Heat map of ISGs generated from a subset extracted from a list of significantly expressed genes in siRNA-treated HeLa cells. Gene expression is depicted by colour intensity. Green denotes upregulation and red downregulation. **b**, RT-qPCR analysis of *IFNB1* mRNA in HeLa cells treated with indicated siRNAs and then 8 h of treatment with vehicle or ABT-737. Data are the mean of two independent experiments. **c**, Western-blot analysis of the cytochrome *c* (cyt *c*) release

into the cytoplasm of HeLa cells treated with vehicle or ABT-737 for 8 h. Subcellular fractionation purity confirmed by relevant markers. Blots are representative of two experiments. **d**, $\log_2(\text{fold change})$ expression levels of ISGs and genes involved in interferon signalling in HepKO versus wild-type female mice. ISG list is based on the Reactome database³³. **e**, Western blot of whole-cell extracts from cells treated with the indicated siRNAs. Blots are representative of two experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | RNA editing of cytoplasmic mtRNA. **a**, RNA editing sites mapped on the RNA transcriptome of SUV3 and PNPase depleted cells are shown. Each dot represents an editing event. Dots on the upper panel denote editing events on the H-strand and dots on the lower panel denote editing events on the L-strand. Triangle denotes single SUV3 editing site. Yellow bars denote the D-loop region. Short red bars denote tRNA genes on the L-strand and green bars denote tRNA genes on the

H-strand. **b**, Frequency of dinucleotide RNA editing sites mapped in the PNPase depleted samples. **c**, RT-qPCR analysis of *IFNB1* mRNA levels in indicated siRNA-treated cells. Data are the mean from two independent experiments. **d**, Western blot of ADAR1, SUV3 and PNPase in cell treated with the indicated siRNAs. Blots are representative of two experiments. For gel source data, see Supplementary Fig. 1.

a**b**

Extended Data Fig. 8 | EMCV infection results in dsRNA accumulation that partially overlaps with mitochondria. **a**, Left, confocal images of EMCV-infected HeLa cell at MOI 1, 6 h after infection, stained with anti-dsRNA (J2) antibody. Mitochondria are stained with MitoTracker Red CMXRos and nucleus with DAPI. Right, line scan RGB profile for the

region of interest (ROI) selected with a white line is shown on the right. Data are representative of two experiments. **b**, Expanded view of the ROI of an EMCV-infected HeLa cell showing colocalization of dsRNA with mitochondria. Image is representative of two experiments. Scale bars, 10 μm.

Extended Data Table 1 | Clinical table of patients carrying PNPT1 mutations

	Patient 1	Patient 2		Patient 3		Patient 4
Nucleotide substitutions	c.407G>A hom	c.208T>C het	c.2137G>T het	c.1495G>C het	c.1519G>T het	c.1160A>G hom ref ¹⁴
Amino acid	p.Arg136His	p.Ser70Pro	p.Asp713Tyr	p.Gly499Arg	p.Ala507Ser	p.Gln387Arg
SIFT	Deleterious (score: 0, median: 3.48)	Tolerated (score: 0.1, median: 3.48)	Deleterious (score: 0.01, median: 3.47)	Deleterious (score: 0.0, median: 3.49)	Deleterious (score: 0.0, median: 3.49)	ref ¹⁴
Mutation Taster	Disease causing (p-value:1)	Disease causing (p-value:0.996)	Disease causing (p-value:1)	Disease causing (p-value:1)	Disease causing (p-value:1)	ref ¹⁴
Polyphen	Probably damaging with score of 1.0 (sensitivity: 0.0; specificity: 1.0)	Probably damaging with score of 0.588 (sensitivity: 0.81; specificity: 0.83)	Probably damaging with score of 0.998 (sensitivity: 0.18; specificity: 0.98)	Probably damaging with score of 0.994 (sensitivity: 0.46; specificity: 0.96)	Probably damaging with score of 0.913 (sensitivity: 0.69; specificity: 0.90)	ref ¹⁴
Splicing prediction				Predicted change at 5'ss (-1G>C): -54.0% MaxEnt: -100.0%; NNSPLICE:49.1%; HSF:-12.9%		
ClinVar					RCV000197606.4	RCV000033022.3
dbSNP	rs746356243	-	-	-	-	-
gnomAD	0.0000244 (6 hets /245,972 alleles)	Not recorded (>230,000 alleles)	Not recorded (>230,000 alleles)	Not recorded (>230,000 alleles)	0.0002241 (62 hets/276,714 alleles)	Not recorded (>230,000 alleles)
Clinical features	Intrauterine growth retardation	Truncal hypotonia		Hypotonia		ref ¹⁴
	Congenital cataract	Nystagmus		Abnormal eye movements		
	Feeding difficulties	Failure to thrive		Sensory neuropathy		
	Deafness	Sensorineural deafness		Deafness		
	Truncal hypotonia	Hypertonia of the lower limbs		Leucodystrophy		
	Nystagmus					
	Death at age 2 years	Alive at age 1 years		Alive at age 7 years		Alive at age 13 years
Brain Imaging	Anomalies of putamen and basal ganglia	Delayed myelination		Delayed myelination; abnormal corpus callosum		ref ¹⁴
Metabolic workup	NMR spectroscopy: lactate peak	NMR spectroscopy: lactate peak		NMR spectroscopy: No lactate elevation		ref ¹⁴
	Hyper-lactatemia	Hyper-lactatemia				
		Hyper-lactatorachia				
	NA	High neopterin in CSF (101 nmol/l)		ND		ND
Interferon score* (blood analysis)	NA	4.138		2.997		2.413
RC analysis	Multiple RC deficiency in muscle and fibroblasts	Multiple RC deficiency in fibroblasts		Normal RC activities in fibroblasts		ref ¹⁴

het = heterozygous; hets = heterozygotes; hom = homozygous; RC = respiratory chain; NA = not available; ND = not determined; *Normal interferon score value is <2.466. SIFT, Sorting intolerant from tolerant (http://sift.bii.a-star.edu.sg/www/SIFT_aligned_seqs_submit.html)

Extended Data Table 2 | Oligonucleotide primers and siRNAs used in the study

Primer (qRT-PCR)	Gene	Sequence (5'-3')
IFN- β forward	<i>IFNB1</i>	ATGACCAACAAGTGTCTCCTCC
IFN- β reverse	<i>IFNB1</i>	GCTCATGGAAAGAGCTGTAGTG
L-mRNA EMCV fwd	L-mRNA EMCV	GCGCACTCTCTCACTTTTGA
L-mRNA EMCV rev	L-mRNA EMCV	TCGAAAACGACTTCCATGTCT
β -actin mRNA forward	<i>ACTB</i>	CTGTGGCATCCACGAAACTA
β -actin mRNA reverse	<i>ACTB</i>	AGTACTTGCGCTCAGGAGGA
COX1 forward	<i>MT-CO1</i>	ACGTTGTAGCCCACTTCCAC
COX1 reverse	<i>MT-CO1</i>	TGGCGTAGGTTTGGTCTAGG
ND5 forward	<i>MT-ND5</i>	TCGAAACCGCAAACATATCA
ND5 reverse	<i>MT-ND5</i>	CAGGCGTTTAATGGGGTTTA
ND6 forward	<i>MT-ND6</i>	CCAATAGGATCCTCCCGAAT
ND6 reverse	<i>MT-ND6</i>	AGGTAGGATTGGTGCTGTGG
<i>CYB</i> forward	<i>MT-CYB</i>	AGACAGTCCCACCCTCACAC
<i>CYB</i> reverse	<i>MT-CYB</i>	GGTGATTCTAGGGGGTTGT
ms Ifit1 forward	<i>Ifit1</i>	CAAGGCAGGTTTCTGAGGAG
ms Ifit1 reverse	<i>Ifit1</i>	GACCTGGTCACCATCAGCAT
ms Gapdh forward	<i>Gapdh</i>	GACTTCAACAGCAACTCCCAC
ms Gapdh reverse	<i>Gapdh</i>	TCCACCACCCTGTTGCTGTA
siRNA	Gene (alternative name)	List of the StealthRNA oligos
siPNPase	<i>PNPT1</i> (PNPase)	HSS131758
siSUV3	<i>SUPV3L1</i> (SUV3)	HSS110378
siMRPP1	<i>TRMT10C</i> (MRPP1)	HSS123550
siMDA5	<i>IFIH1</i> (MDA5)	HSS127414
siTLR3	<i>TLR3</i>	HSS110815
siDDX28	<i>DDX28</i>	HSS125053
siEXOG	<i>EXOG</i>	HSS115058
siMGME1	<i>MGME1</i>	HSS132389
siFEN1	<i>FEN1</i>	HSS103629
siRNA	Gene	siRNA Forward Strand
siCntrl	Luciferase	GAUUAUGUCCGUUAUGUAUU
siRIG-I	<i>DDX58</i> (RIG-I)	ACGGAUUAGCGACAAAUUUAA
siADAR1	<i>ADAR</i> (ADAR1)	sc-37657
ON-TARGETplus Human BAK	<i>BAK</i>	L-003305-00-0005
ON-TARGETplus Human BAX	<i>BAX</i>	L-003308-01-0005
ON-TARGETplus Human MAVS	<i>MAVS</i>	L-024237-00-0005
ON-TARGETplus Human SUV3 *	<i>SUPV3L1</i>	L-017841-01-0020
ON-TARGETplus Human PNPase *	<i>PNPT1</i>	L-019454-01-0020

CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window

Shakked O. Halperin^{1,2,3}, Connor J. Tou¹, Eric B. Wong¹, Cyrus Modavi^{1,2}, David V. Schaffer^{1,3,4,5,6*} & John E. Dueber^{1,3,7*}

The capacity to diversify genetic codes advances our ability to understand and engineer biological systems^{1,2}. A method for continuously diversifying user-defined regions of a genome would enable forward genetic approaches in systems that are not amenable to efficient homology-directed oligonucleotide integration. It would also facilitate the rapid evolution of biotechnologically useful phenotypes through accelerated and parallelized rounds of mutagenesis and selection, as well as cell-lineage tracking through barcode mutagenesis. Here we present EvolvR, a system that can continuously diversify all nucleotides within a tunable window length at user-defined loci. This is achieved by directly generating mutations using engineered DNA polymerases targeted to loci via CRISPR-guided nickases. We identified nickase and polymerase variants that offer a range of targeted mutation rates that are up to 7,770,000-fold greater than rates seen in wild-type cells, and editing windows with lengths of up to 350 nucleotides. We used EvolvR to identify novel ribosomal mutations that confer resistance to the antibiotic spectinomycin. Our results demonstrate that CRISPR-guided DNA polymerases enable multiplexed and continuous diversification of user-defined genomic loci, which will be useful for a broad range of basic and biotechnological applications.

Natural biological systems evolve astounding functionality by sampling immense genetic diversity under selective pressures. Forward genetics emulates this process to help us understand naturally evolved biological phenomena and to direct the evolution of biotechnologically useful material by applying an artificial selection pressure or screen to libraries of genetic variants. New forward genetic approaches would be enabled by a targeted mutator capable of continuously diversifying all nucleotides within user-defined regions of a genome. However, previous targeted continuous-diversification techniques are confined to either evolving specific loci within particular cells under stringent culture conditions^{2,3} or mutating particular types of nucleotides in a narrow, user-defined window^{4,5} (Extended Data Fig. 1). Conversely, current techniques capable of diversifying all nucleotides within user-defined loci remain discrete owing to their requirement for efficient integration of oligonucleotide libraries at the target site^{6,7}. Therefore, no method currently exists to continuously diversify all nucleotides within user-defined regions of a genome (Extended Data Table 1).

DNA polymerases have the capacity to create all 12 substitutions, as well as deletions^{8,9}. These enzymes vary in processivity (average number of nucleotides incorporated after each binding event), fidelity (misincorporation rate) and substitution bias (nucleotide bias during misincorporation). In particular, nick-translating DNA polymerases are able to initiate synthesis from a single-stranded break in double-stranded DNA while displacing the downstream nucleotides, and their flap endonuclease domain subsequently degrades the displaced nucleotides, leaving a ligatable nick. We hypothesized that recruiting an error-prone, nick-translating DNA polymerase with a nicking variant of Cas9¹⁰ (nCas9) could offer an ideal targeted mutagenesis tool that is

independent of homology-directed repair, and which we term EvolvR (Fig. 1a). The specificity of the polymerase initiation site created by the nCas9 specifies the start site of the editing window, and the mutagenesis window length, mutation rate and substitution bias are controlled by the processivity, fidelity and misincorporation bias of the polymerase variant, respectively.

In the initial design, nCas9 (*Streptococcus pyogenes* Cas9 containing a D10A mutation) was fused to the N terminus of a fidelity-reduced variant of *Escherichia coli* DNA polymerase I (PolI) harbouring the mutations D424A, I709N and A759R (PolI3M)³. A plasmid (pEvolvR) expressing the nCas9–PolI3M and a guide RNA (gRNA) in *E. coli* was tested for its ability to mutate a second plasmid targeted by the gRNA over 24 h of propagation. High-throughput targeted amplicon sequencing revealed that the target plasmid accrued substitutions in an approximately 17-nucleotide window 3' of the nick site (Fig. 1b), consistent with the established 15–20 nucleotide processivity of PolI¹¹. Although the sequencing results are probably under-sampling the total diversity generated, we observed substitutions of all four nucleotide types (Fig. 1c). The presence of low-frequency substitutions 5' of the nick site may be due to endogenous 3'-to-5' exonucleases removing a few nucleotides 5' of the nick prior to the polymerase initiating synthesis. Controls expressing an unfused nCas9 and PolI3M with the on-target guide only yielded one low-frequency substitution, whereas expressing nCas9–PolI3M with an off-target guide, as well as expressing nCas9 alone, did not yield any substitutions at a frequency above our detection threshold.

To sensitively quantify the mutation rate and mutagenesis window length of EvolvR variants, we designed a fluctuation analysis¹². For this assay, the pEvolvR plasmid was co-transformed into *E. coli* with a plasmid (pTarget) containing the *aadA* spectinomycin resistance gene disabled by a nonsense mutation (Fig. 1d). After 16 h of growth, the cultures were plated on spectinomycin and the mutation rates were determined from the number of resistant colony-forming units (CFUs). As shown in Fig. 1e, fluctuation analysis estimated the mutation rate of wild-type *E. coli* to be approximately 10^{-10} mutations per nucleotide per generation, similar to previously reported values¹³. The global mutation rate (the mutation rate of the untargeted genome in cells expressing EvolvR) was determined by measuring the spectinomycin-resistance reversion rate of cells carrying a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome¹⁴, whereas the targeted mutation rate was determined with a gRNA nicking 11 nucleotides 5' of the nonsense mutation in pTarget. Expressing nCas9–PolI3M markedly increased the mutation rate at the targeted locus 24,500-fold over the wild type while increasing the global mutation rate 120-fold over the wild type (Fig. 1e), a global mutation rate comparable to that of previous targeted mutagenesis techniques in *E. coli*^{3,6}. By comparison, expressing nCas9 and PolI3M as separate proteins, PolI3M alone, nCas9 alone or a catalytically inactive Cas9 (dCas9) fused to PolI3M, showed significantly lower targeted mutation rates ($P < 0.0001$;

¹Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA. ²University of California, Berkeley–University of California, San Francisco Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA, USA. ³Innovative Genomics Institute, University of California Berkeley and San Francisco, Berkeley, CA, USA. ⁴Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, CA, USA. ⁵Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. ⁶Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. ⁷Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. *e-mail: schaffer@berkeley.edu; jdueber@berkeley.edu

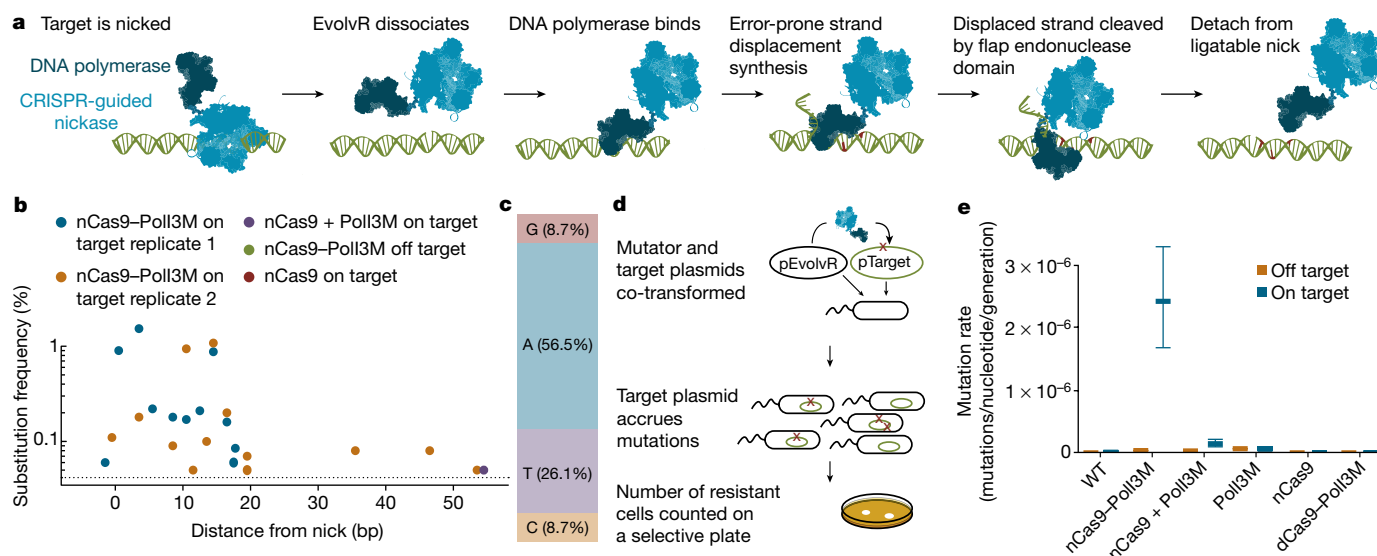


Fig. 1 | EvolvR enables targeted mutagenesis. **a**, The EvolvR system consists of a CRISPR-guided nickase that nicks the target locus and a fused DNA polymerase that performs error-prone nick translation. **b**, High-throughput sequencing shows that fusing nCas9 to PolI3M resulted in substitutions across an approximately 17-nucleotide window 3' from the nick. Expressing nCas9–PolI3M with an off-target guide did not show substitutions at a frequency above our detection threshold (dotted line, see Methods 'High-throughput sequencing data analysis'), whereas an unfused nCas9 and PolI3M yielded only one substitution and at low-frequency. **c**, Distribution of the substituted nucleotides; all four nucleotides were substituted by nCas9–PolI3M. **d**, Schematic of the

fluctuation analysis workflow used to sensitively quantify targeted and global mutation rates. **e**, The global and targeted mutation rates of wild-type (WT) *E. coli*, nCas9–PolI3M, unfused nCas9 and PolI3M, PolI3M alone, and nCas9 alone were determined by fluctuation analysis. For all figures, 'on target' mutation rates were determined by expressing a gRNA that nicks 11 nucleotides 5' of the nonsense mutation unless labelled otherwise, whereas the 'off target' mutation rates were determined by expressing a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome. Data are the mean of ten biologically independent samples and the error bars indicate 95% confidence intervals. Mutation rates throughout are mutations per nucleotide per generation.

two-sided student's *t*-test). These results suggest that both PolI3M and the nick created by nCas9 are essential for EvolvR-mediated mutagenesis. Expressing nCas9 and PolI3M as separate proteins or PolI3M alone showed a 98-fold or 554-fold increase in global mutation rates compared to wild-type *E. coli*, respectively. Finally, by replacing the D10A nCas9—which nicks the strand complementary to the gRNA—with the H840A nickase, which nicks the strand non-complementary to the gRNA, we found that the direction of EvolvR-mediated mutagenesis relative to the target site of the gRNA is dependent on which strand is nicked (Extended Data Fig. 2).

We hypothesized that the targeted mutation rate could be further increased by promoting the dissociation of nCas9 from DNA after nicking the target locus. Therefore, three mutations (K848A, K1003A, R1060A) that have previously been suggested to lower the non-specific DNA affinity of Cas9¹⁵ were introduced into the fused nCas9. The resulting enhanced nCas9 (enCas9) fused to PolI3M increased the global mutation rate 223-fold compared to wild-type cells (1.9-fold greater than nCas9–PolI3M), yet elevated the mutation rate at the targeted locus by 212,000-fold (8.7-fold greater than nCas9–PolI3M) (Fig. 2a).

PolI3M was initially chosen because it was the most error-prone variant of PolI previously characterized. However, the modularity of EvolvR enables tuning of the mutation rate by using polymerases with different fidelities. First, we confirmed that the fidelity of the polymerase determines mutation rates by comparing enCas9 fused to PolI variants, in decreasing order of fidelity: PolI1M (D424A), PolI2M (D424A, I709N), and PolI3M (D424A, I709N, A759R) (Fig. 2b). Next, to further increase the targeted mutation rate of EvolvR, we screened several additional mutations previously shown to individually decrease wild-type PolI fidelity^{3,16,17} (Fig. 3c). Although several of the additional mutations yielded low-activity variants, PolI3M with the additional mutations F742Y and P796H (PolI5M) displayed a mutation rate one nucleotide from the nick that was 7,770,000-fold higher than wild-type cells, and 33-fold higher than PolI3M, making it the most error-prone PolI mutant ever reported. Notably, enCas9–PolI5M did not exhibit either a higher global rate of mutagenesis than enCas9–PolI3M or higher

mutation rates than enCas9–PolI3M 11 nucleotides from the nick (Extended Data Fig. 3).

A more processive DNA polymerase could potentially increase the length of the editing window, so PolI5M was exchanged for the more processive bacteriophage Phi29 DNA polymerase (Phi29). Expression of Phi29 variants with previously reported fidelity-reducing and thermostabilizing mutations in combination with the Phi29 single-stranded binding protein showed targeted mutagenesis 56 and 347 nucleotides from the nick site (Extended Data Fig. 4). However, the mutation rate at these distances was not nearly as high as that achieved with PolI3M at shorter distances.

An alternative method to increase the length of the editing window and retain high mutation rates would be to increase the processivity of PolI. Previous work has shown that inserting the thioredoxin-binding domain (TBD) of bacteriophage T7 DNA polymerase into the thumb domain of PolI increases the processivity of the polymerase in the presence of thioredoxin from *E. coli*¹⁸. Figure 2d shows that whereas the original enCas9–PolI3M did not show a difference between global and targeted mutation rates 56 nucleotides from the nick, incorporation of the TBD into the PolI3M EvolvR gene (enCas9–PolI3M–TBD) produced a 555-fold increase over the global mutation rate at this range. Leveraging this increased editing window length, we targeted enCas9–PolI3M–TBD to a plasmid (pTarget2) containing two nonsense mutations (11 and 37 nucleotides from the nick) in the antibiotic-resistance gene, and thereby showed the ability of EvolvR to generate combinations of multiple mutations with a single gRNA (Fig. 2e).

We hypothesized that unintended translation products consisting of functional DNA polymerase not fused to a functional CRISPR-guided protein contributed to undesirable off-target mutagenesis. Therefore, we codon-optimized the EvolvR coding sequence (enCas9–PolI3M–TBD–CO) to remove three strong internal ribosome binding sites identified using the RBS Calculator¹⁹. We found that the off-target mutation rate decreased 4.14-fold when expressing enCas9–PolI3M–TBD–CO compared to enCas9–PolI3M–TBD while the on-target mutation rate only decreased 1.23-fold (Extended Data Fig. 5).

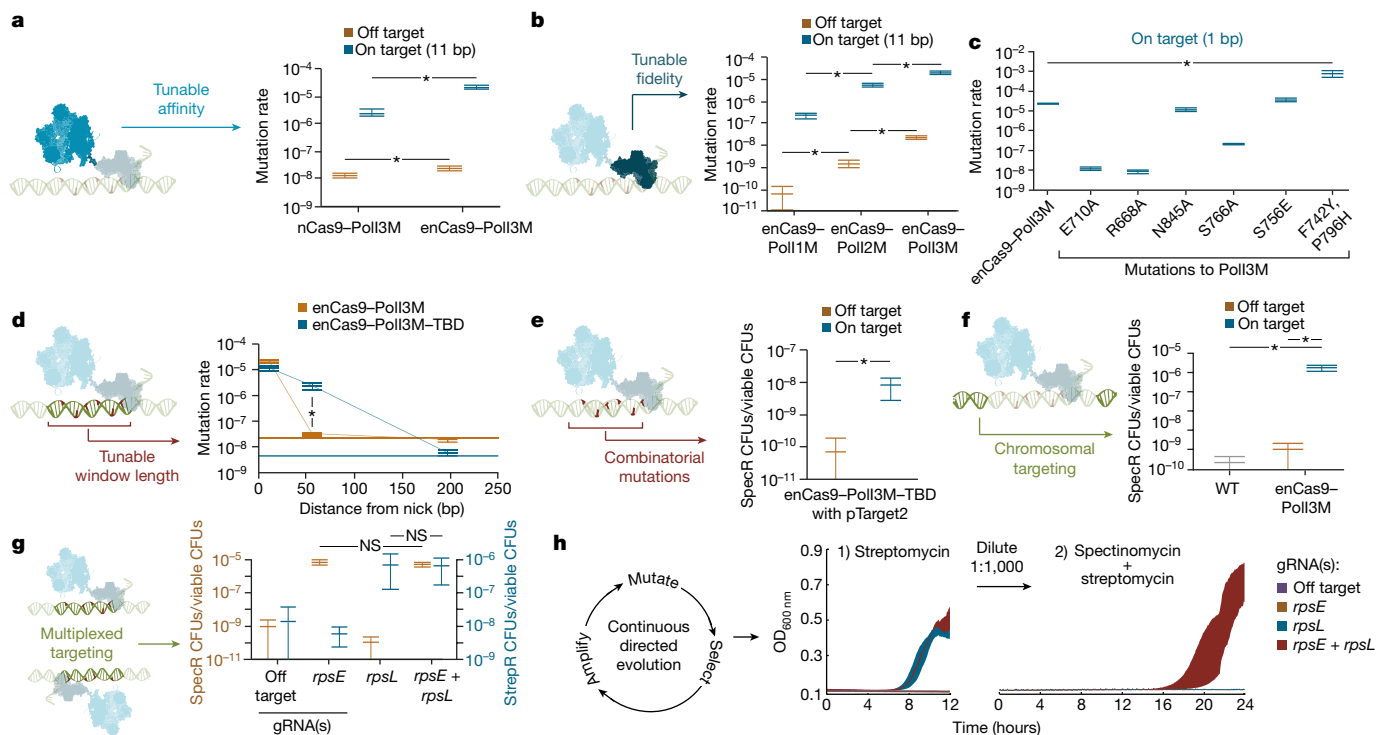


Fig. 2 | EvolvR provides tunable mutation rates and mutagenesis-window lengths, combinatorial mutations, multiplexed targeting and continuous diversification of genomic loci. **a**, Introducing mutations suggested to lower non-specific DNA affinity into the fused nCas9 (producing enCas9)¹⁵ increased the global mutation rate 223-fold compared to the wild-type mutation rate (enCas9–PolI3M 1.9-fold greater than nCas9–PolI3M), and increased the targeted mutation rate by 212,038-fold over the wild type (enCas9–PolI3M 8.7-fold greater than nCas9–PolI3M). **b**, Mutagenesis rates were dependent on the fidelity of the polymerase. PolI with a D424A mutation (PolI1M) was less mutagenic than PolI with both D424A and I709N mutations (PolI2M), and PolI3M (D424A, I709N, A759R) was the most mutagenic. **c**, Screening mutations in PolI3M previously shown to decrease wild-type PolI fidelity revealed that PolI3M with additional mutations F742Y and P796H (PolI5M) had a mutation rate 7,770,000-fold greater than wild-type cells one nucleotide from the nick. **d**, The editing-window length was increased by incorporating TBD into PolI3M. enCas9–PolI3M–TBD provided a targeted mutation rate 56 nucleotides from the nick that was 555-fold above the global mutation rate, whereas enCas9–PolI3M showed no targeted mutagenesis 56 nucleotides from the nick. **e**, enCas9–PolI3M–

The ability to couple EvolvR-mediated mutagenesis to a genetic screen of a non-selectable phenotype would considerably broaden the utility of EvolvR. We found that after targeting EvolvR to a plasmid containing a green fluorescent protein (GFP) cassette with an early termination codon, 0.06% and 0.07% of the population expressed GFP, whereas no cells expressed GFP when an off-target gRNA was used (Extended Data Fig. 6). Importantly, EvolvR also showed the capacity to diversify chromosomal loci by increasing the fraction of the population resistant to spectinomycin approximately 16,000-fold after targeting enCas9–PolI3M to the endogenous ribosomal protein subunit 5 gene of *E. coli* (*rpsE*), which has mutations that are known to confer resistance to spectinomycin²⁰ (Fig. 2f).

Next, we tested whether EvolvR avoids the toxicity associated with non-targeted mutagenesis systems²¹. We found that, unlike two previously developed non-targeted continuous-mutagenesis systems, EvolvR does not impede cell viability or growth rate (Extended Data Fig. 7a, b). Additionally, targeting EvolvR to the *rpsE* gene evolved more spectinomycin-resistant CFUs per ml compared to these previous non-targeted mutagenesis systems (Extended Data Fig. 7c). Finally, targeting EvolvR to a GFP cassette containing a nonsense mutation resulted in 28 times more GFP-positive cells than when using the most

TBD targeted to a plasmid containing two nonsense mutations in the spectinomycin resistance gene (pTarget2) showed that EvolvR is able to generate combinations of multiple mutations. **f**, enCas9–PolI3M targeted to *E. coli rpsE* generated approximately 16,000-fold more spectinomycin-resistant CFUs (SpecR CFUs) than when targeted to the *dbpA* locus. **g**, enCas9–PolI3M–TBD targeted to *rpsL* increased the rate of acquiring streptomycin resistance without increasing the rate of acquiring spectinomycin resistance. Coexpression of both *rpsL* and *rpsE* gRNAs increased both spectinomycin and streptomycin resistant CFUs. **h**, Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in streptomycin-supplemented medium, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not. After back-dilution into spectinomycin- and streptomycin-supplemented media, only cultures expressing both *rpsE* and *rpsL* gRNAs grew. Mutation rate data are mean ± 95% confidence intervals from ten biologically independent samples. Resistant CFUs/viable CFUs data are mean ± s.d. from ten biologically independent samples. **P* < 0.05; two-sided student's *t*-test. In **h**, the shaded region of OD_{600 nm} indicates mean ± s.d. from three biologically independent samples.

recently developed non-targeted continuous mutagenesis technique (Extended Data Fig. 8).

EvolvR could enable simultaneous diversification of distant genomic loci through coexpression of multiple gRNAs. Expression of a gRNA targeting enCas9–PolI3M–TBD to *rpsL*, a ribosomal protein subunit gene capable of acquiring mutations that confer streptomycin resistance²², increased the rate of acquiring streptomycin resistance compared to wild-type cells, without altering sensitivity to spectinomycin (Fig. 2g). By comparison, coexpression of the gRNAs targeting *rpsE* and *rpsL* generated approximately the same number of respective spectinomycin- and streptomycin-resistant CFUs as observed for individual expression of the *rpsE* gRNA (*P* = 0.0752; two-sided Student's *t*-test) and *rpsL* gRNA (*P* = 0.885; two-sided Student's *t*-test). This capacity to simultaneously diversify multiple loci will be useful for identifying epistatic interactions. We also note that the expression of two gRNAs that nick separate strands at genomic loci separated by 100 bp was lethal, whereas nicking the same strand at this 100-bp distance was not lethal. Therefore, if multiple gRNAs are to be used to increase the length of the target region, we recommend targeting the same strand.

To evolve resistance to both spectinomycin and streptomycin, we used the continuous diversity generation of EvolvR for continuous

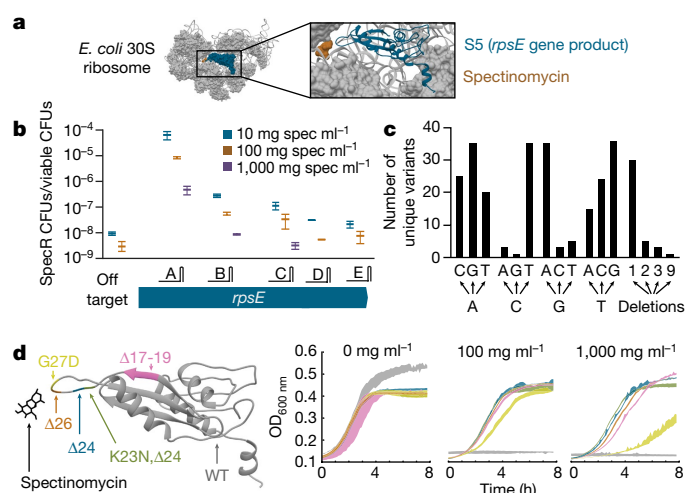


Fig. 3 | EvolvR identified novel mutations to the *E. coli* *rpsE* gene that confer spectinomycin resistance. **a**, Spectinomycin inhibits protein synthesis through interactions with the 30S ribosome. **b**, enCas9–PolI3M–TBD targeted to different parts of the endogenous *rpsE* gene with five gRNAs showed higher rates of spectinomycin (spec) resistance than targeting *dbpA* (off-target). Data are mean \pm s.d. from three biologically independent samples. **c**, After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions as well as deletions were generated. **d**, Left, five mutations not previously described as conferring spectinomycin resistance were regenerated in a new strain of *E. coli* (RE1000). Right, growth curves in varying concentrations of spectinomycin confirmed that the mutations provide spectinomycin resistance. Shaded area represents mean \pm s.d. from three biologically independent samples.

directed evolution (in which mutagenesis, selection and amplification occur simultaneously) to allow adaptation to modulated selection pressures with minimal researcher intervention. Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in liquid medium supplemented with streptomycin, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not (Fig. 2h). After the cultures were diluted 1,000-fold into liquid medium supplemented with both spectinomycin and streptomycin, only cultures expressing both *rpsE* and *rpsL* gRNAs grew.

The clinical utility of spectinomycin as a broad-spectrum antibiotic has motivated previous efforts to characterize genomic mutations conferring spectinomycin resistance²³. We used the capacity of EvolvR to diversify the genomic *rpsE* gene to identify novel mutations that confer spectinomycin resistance by disrupting the spectinomycin-binding pocket of the 30S ribosome (Fig. 3a). First, we targeted enCas9–PolI3M–TBD to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair within the 504-bp *rpsE* coding sequence (Extended Data Fig. 9a). Then, we challenged the cell populations for growth on agar plates supplemented with varying concentrations of spectinomycin and observed that resistance was highest with the gRNAs targeted to the domain of the ribosomal subunit protein that is proposed to interact with spectinomycin (Fig. 3b). After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions, as well as deletions, were generated (Fig. 3c). For functional analysis, we introduced five of the candidate mutations not previously described as providing spectinomycin resistance into a different strain of *E. coli* (RE1000) using oligonucleotide-mediated recombination. Growth curves in varying concentrations of spectinomycin confirmed that each of the five mutations (Δ 17–19; K23N, Δ 24; Δ 24; Δ 26; G27D) provided varying levels of spectinomycin resistance, but reduced fitness in the absence of selection (Fig. 3d). On the basis of these mutations, we hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket confer resistance to spectinomycin by removing a hydrogen bond that stabilizes the interaction of

spectinomycin with the ribosome. Therefore we tested an array of deletions that we predicted would move Lys26 and discovered additional novel mutations that confer spectinomycin resistance (Extended Data Fig. 9b, c). This rapid method for discovering genotypes conferring antibiotic resistance will be generally useful for improving the effective use of antibiotics.

EvolvR offers the first example of continuous targeted diversification of all nucleotides at user-defined loci, which will be useful for evolving protein structure and function, mapping protein–protein and protein–drug interactions, investigating the non-coding genome, engineering industrially relevant microbes and tracking the lineage of cell populations that cannot tolerate double-stranded breaks²⁴. As a guiding principle for using this tool, our data suggest that 1 μ l saturated *E. coli* culture expressing enCas9–PolI3M–TBD for 16 h contains all single substitutions in the 60-nucleotide window with more than tenfold coverage. Future work towards adapting EvolvR for use in cells possessing low transformation efficiency, as well as increasing the mutation rate and window length of EvolvR mutagenesis, would enable new forward genetic applications.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0384-8>.

Received: 1 December 2017; Accepted: 19 June 2018;

Published online 1 August 2018.

- Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable continuous evolution of genes at mutation rates above genomic error thresholds. Preprint at <https://www.biorxiv.org/content/early/2018/05/03/313338> (2018).
- Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
- Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl Acad. Sci. USA* **100**, 9727–9732 (2003).
- Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
- Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
- Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
- Costantino, N. & Court, D. L. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl Acad. Sci. USA* **100**, 15748–15753 (2003).
- Troll, C., Alexander, D., Allen, J., Marquette, J. & Camps, M. Mutagenesis and functional selection protocols for directed evolution of proteins in *E. coli*. *J. Vis. Exp.* **49**, e2505 (2011).
- de Boer, J. G. & Ripley, L. S. An in vitro assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. *Genetics* **118**, 181–191 (1988).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Bambara, R. A., Uyemura, D. & Choi, T. On the processive mechanism of *Escherichia coli* DNA polymerase I. Quantitative assessment of processivity. *J. Biol. Chem.* **253**, 413–423 (1978).
- Sarkar, S., Ma, W. T. & Sandri, G. H. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173–179 (1992).
- Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164 (1991).
- Jagessar, K. L. & Jain, C. Functional and molecular analysis of *Escherichia coli* strains lacking multiple DEAD-box helicases. *RNA* **16**, 1386–1392 (2010).
- Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
- Minnick, D. T. et al. Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067–3075 (1999).
- Loh, E., Salk, J. J. & Loeb, L. A. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc. Natl Acad. Sci. USA* **107**, 1154–1159 (2010).
- Wang, Y. et al. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32**, 1197–1207 (2004).
- Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42 (2011).

20. Funatsu, G., Schiltz, E. & Wittmann, H. G. Ribosomal proteins. XXVII. Localization of the amino acid exchanges in protein S5 from two *Escherichia coli* mutants resistant to spectinomycin. *Mol. Gen. Genet.* **114**, 106–111 (1972).
21. Zheng, X., Xing, X.-H. & Zhang, C. Targeted mutagenesis: a sniper-like diversity generator in microbial engineering. *Synth. Syst. Biotechnol.* **2**, 75–86 (2017).
22. Timms, A. R., Steingrimsdottir, H., Lehmann, A. R. & Bridges, B. A. Mutant sequences in the *rpsL* gene of *Escherichia coli* B/r: mechanistic implications for spontaneous and ultraviolet light mutagenesis. *Mol. Gen. Genet.* **232**, 89–96 (1992).
23. Brocklehurst, P. Antibiotics for gonorrhoea in pregnancy. *Cochrane Database of Systematic Reviews* **2**, CD000098 <https://doi.org/10.1002/14651858.CD000098> (2002).
24. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).

Acknowledgements We thank S. McDevitt at the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for assistance with high-throughput sequencing, the Arkin laboratory for supplying *E. coli* strain RE1000, W. DeLoache for helping edit our manuscript, and the Innovative Genomics Institute for funding.

Author contributions S.O.H. conceived of all designs, designed the study, contributed to the execution of all experiments, analysed all of the data and

wrote the manuscript; C.J.T. contributed to plasmid construction and assay execution for fluctuation analyses, and spectinomycin-resistance mutation identification; E.B.W. contributed to plasmid construction and assay execution for Poll3M mutant screening, Phi29 screening, multiplexing and spectinomycin-resistance identification; S.O.H., C.M., D.V.S. and J.E.D. contributed to assay design. The manuscript was read, edited and approved by all authors.

Competing interests The Regents of the University of California have filed a provisional patent application (62/662,043 and 62/556,127) related to the technology described in this work to the United States Patent and Trademark Office; S.O.H. is listed as the inventor.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0384-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0384-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.V.S. and J.E.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Plasmid construction. All plasmids were constructed using a modular Golden Gate strategy. pEvolVR consisted of EvolvR and gRNA expression cassettes, a pBR322 origin of replication and a kanamycin resistance cassette. pTarget consisted of a p15a origin of replication carrying both a functional trimethoprim resistance cassette for selection and a disabled spectinomycin resistance gene (*aadA*) harbouring a L106X nonsense mutation. pTarget2 is identical to pTarget except that the *aadA* gene now carried both Q98X and L106X mutations. The full plasmid sequences are provided in Supplementary Table 1.

High-throughput sequencing of pTarget sample preparation. A pTarget and pEvolVR plasmid were cotransformed into 50 µl chemically competent TG1 *E. coli* prepared by a TSS/KCM method. Cells were allowed to recover in the TSS/LB solution for 1 h, before 4 µl transformation mix was inoculated into 2 ml LB containing 25 µg/ml kanamycin and 15 µg/ml trimethoprim. The cultures were grown for 24 h at 37 °C while shaking at 750 rpm. A 1.5-ml sample of each culture was miniprep using a Zippy Plasmid Prep kit (Zymo Research).

The oligonucleotides pTarget-F and pTarget-R were used to amplify the target region in a 20-cycle PCR reaction using 100 ng miniprep DNA as the template. A second PCR reaction added Illumina sequencing adapters and indices to the previous PCR product over 10 thermocycles. A Qubit fluorimeter was used to quantify the DNA before pooling samples. The sample pool was submitted to the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for quality control and sequencing. Quality control consisted of fragment analysis (Advanced Analytical) and concentration measurement of the sequenceable fraction by quantitative PCR (Kapa Biosystems). The pooled sample was mixed with Illumina PhiX sequencing control library at 10% molarity, diluted to 14 pM, denatured, and run on an Illumina MiSeq using a 150-bp paired-end read MiSeq Reagent Kit v2. Resulting basecall files were converted into demultiplexed fastq format using Illumina bcl2fastq v.2.17.

High-throughput sequencing data analysis. Perfectly complementary paired reads were filtered, and the five randomized nucleotides, amplification primer sequences, and first and last five nucleotides were trimmed using a custom Python script. Bwa and samtools were used to generate alignment files using the wild-type *aadA* gene sequence as a reference. VarScan2 was used for variant calling with the parameters: min-coverage 1; min-reads 2; variants 1; min-var.-freq 0.0005; p-value 0.99²⁵. The limit of detection was determined by sequencing a culture transformed with an empty vector as a control. The highest frequency variant was 0.04% so all substitutions with a frequency under 0.05% were discarded.

Fluctuation analysis assay. A 50-µl sample of chemically competent TG1 *E. coli* were cotransformed with pEvolVR and pTarget or pTarget2. After 1 h of recovery at 37 °C, 4 µl was inoculated into a 1.996 ml LB containing 25 µg/ml kanamycin and 15 µg/ml trimethoprim. After shaking at 37 °C for 16 h, 1 ml and 1 µl culture were plated on separate LB agar plates containing 50 µg/ml spectinomycin. For viable CFU counting, 300 µl of 1:50,000,000 diluted culture was plated on LB agar plates. After 24 h of incubation at 37 °C, spectinomycin-resistant CFUs and viable CFUs were counted. Ten replicates were used for each condition.

Calculation of mutation rate and statistics. The Ma–Sandri–Sarkar Maximum Likelihood Estimator was used to determine mutation rates as it is the most accurate and valid for all mutation rates¹². FalcOR was used to calculate the mutation rates by inputting the viable and resistant CFU counts for the ten replicates²⁶. A two-tailed Student's *t*-test was carried out to determine *P* values as previously described²⁷.

Fluorescence-activated cell sorting of EvolvR libraries. pEvolVR expressing either an on- or off-target gRNA was cotransformed with pTarget-GFP* and shaken at 37 °C for 24 h. For each sample, the GFP positive fraction of a million events was sorted with a Cell Sorter SH800 (Sony) using a 488-nm laser and a 525/50-nm emission filter.

Continuous evolution of *E. coli* resistant to both spectinomycin and streptomycin. pEvolVR expressing enCas9–PolI3M–TBD and either the off-target gRNA (targeting *dbpA*), *rpsL* gRNA, *rpsE* gRNA or both *rpsL* and *rpsE* gRNAs was transformed into TG1 *E. coli* as previously described. After recovering for one hour, 4 µl of transformation mix was inoculated into 2 ml of LB supplemented with 25 µg/ml kanamycin and cultures were propagated over 16 h at 37 °C. For each culture 2 µl of culture was re-inoculated into 198 µl of LB supplemented with 50 µg/ml of streptomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 12 h of growth at 37 °C. Each well was then diluted 1,000-fold into LB supplemented with 50 µg/ml of streptomycin and 25 µg/ml of

spectinomycin and the optical density of 200 µl of culture was again measured with a Tecan M1000 Pro spectrophotometer over 24 h of growth at 37 °C. Three biological replicates for each gRNA were characterized.

High-throughput sequencing of spectinomycin resistant *E. coli*. A pEvolVR plasmid expressing enCas9–PolI3M–TBD with *rpsE* gRNA A, B, C, D or E was transformed into chemically competent TG1 *E. coli*. Cells were allowed to recover for 1 h before inoculating 4 µl transformation mix into 1.996 ml LB supplemented with 25 µg/ml kanamycin. The cultures were grown for 16 h at 37 °C while shaking. One millilitre and one microlitre of each culture were plated on separate LB agar plates containing 10, 100, or 1,000 µg/ml spectinomycin. Resistant CFUs were counted in the same manner as the fluctuation assays. The colonies from each plate were then pooled into separate cultures containing 2 ml of LB supplemented with 50 µg/ml spectinomycin and grown for 16 h at 37 °C. Genomic DNA was purified using the Wizard Genomic DNA Purification Kit (Promega). One hundred nanograms of purified genome was then processed and sequenced in the same manner as already described for the sequencing analysis of pTarget, with the one alteration that the oligonucleotides *rpsE*-F and *rpsE*-R were used for the first round of PCR.

Oligonucleotide recombination. Re-introduction of *rpsE* mutations was performed using RE1000 *E. coli* (MG1655 λ-Red::bioA/bioB ilvG+ pTet2:gam-bet-exo-dam pN25:tetR dnaG.Q576A lacIQ1 Pcp8-araE ΔaraBAD pConst-araC ΔrecJ ΔxonA) developed for recombineering. Electro-competent cells were prepared fresh from overnight cultures of bacteria. The saturated culture was back-diluted 1:70 into 5 ml LB with 100 ng/µl anhydrous tetracycline and shaken at 37 °C until the optical density reached 0.5. Cultures were then transferred to an ice-water bath and swirled for approximately 30 s before being chilled on ice for 10 min. Chilled cultures were centrifuged at 9,800g for 1 min. The supernatant was aspirated and the pellet was resuspended in 1 ml ice-chilled 10% glycerol. Washing with glycerol was repeated twice. The final pellet was resuspended in 70 µl chilled 10% glycerol for each transformation. 1 µg of oligonucleotide was electroporated into the cells. The cells were recovered for 1 h at 37 °C in 1 ml LB and streaked out on LB agar plates containing 50 µg/ml spectinomycin. Successful recombination was verified by Sanger sequencing a PCR amplification of the genomic *rpsE* gene.

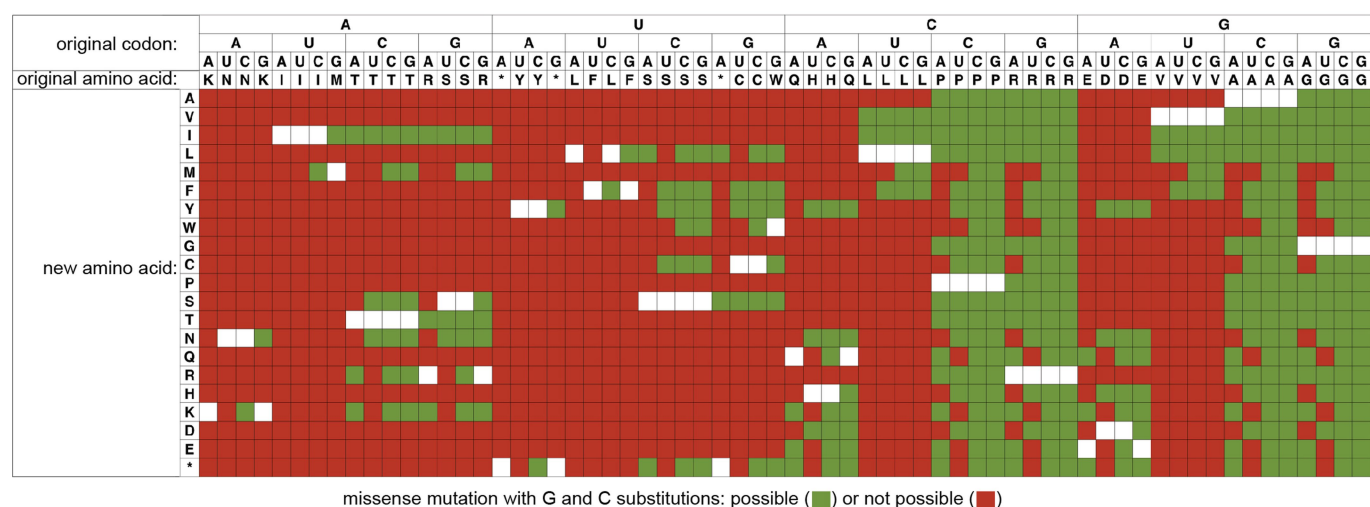
Characterization of spectinomycin resistance. Single colonies of sequence-verified *rpsE* mutants were grown overnight in LB media and then back-diluted 1:200 into LB containing 0, 100 or 1,000 µg/ml spectinomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 8 h of growth at 37 °C. Three biological replicates of each mutant at each spectinomycin concentration were characterized.

Code availability. The code that support the findings of this study are available at <https://github.com/sohnx8/EvolvR>.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

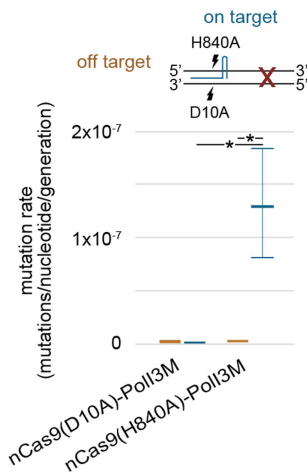
Data availability. The data that support the findings of this study are available from the corresponding authors upon request. High-throughput sequencing data have been deposited as a NCBI BioProject under accession number PRJNA472658. Plasmids encoding enCas9–PolI3M–TBD and enCas9–PolI5M are available from Addgene (plasmids 113077 and 113078).

25. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
26. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. Fluctuation analysis Calculator: a web tool for the determination of mutation rate using Luria–Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
27. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20**, 4–17 (2000).
28. Truniger, V., Lázaro, J. M., de Vega, M., Blanco, L. & Salas, M. phi 29 DNA polymerase residue Leu384, highly conserved in motif B of eukaryotic type DNA replicases, is involved in nucleotide insertion fidelity. *J. Biol. Chem.* **278**, 33482–33491 (2003).
29. de Vega, M., Lázaro, J. M., Salas, M. & Blanco, L. Primer-terminus stabilization at the 3'-5' exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. *EMBO J.* **15**, 1182–1192 (1996).
30. Ducani, C., Bernardinelli, G. & Högborg, B. Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Res.* **42**, 10596–10604 (2014).
31. Povilaitis, T., Alzbutas, G., Sukackaitė, R., Siurkus, J. & Skirgaila, R. *In vitro* evolution of phi29 DNA polymerase using isothermal compartmentalized self replication technique. *Protein Eng. Des. Sel.* **29**, 617–628 (2016).
32. Badran, A. H. & Liu, D. R. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
33. Greener, A., Callahan, M. & Jernseth, B. An efficient random mutagenesis technique using an *E. coli* mutator strain. *Mol. Biotechnol.* **7**, 189–195 (1997).

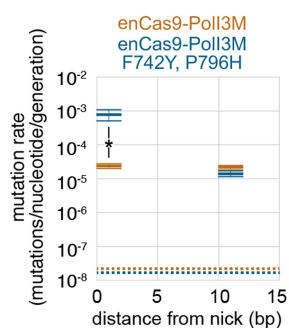


Extended Data Fig. 1 | Bias of cytidine deaminase-mediated targeted diversification. Previous tools enabling diversification of user-defined loci by substituting cytosines and guanines limit the protein coding space that can be explored^{4,5}. This chart shows which amino acids can (green) and

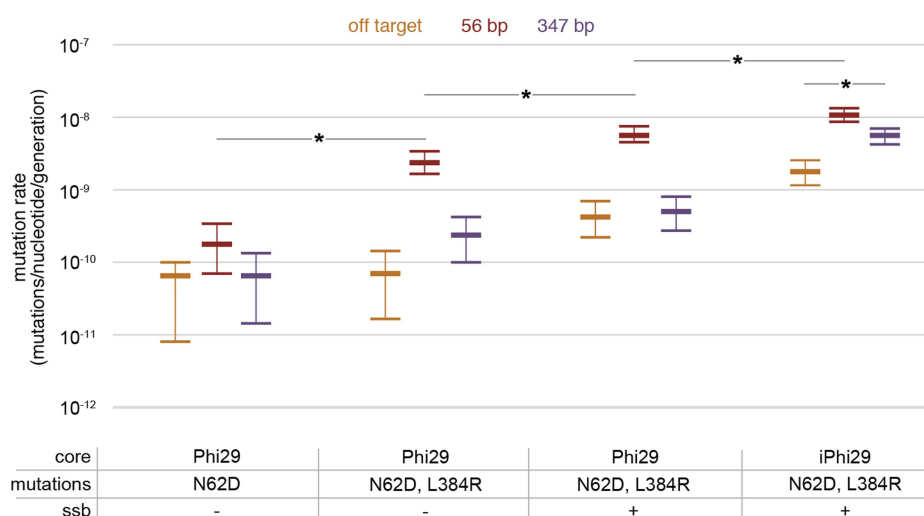
cannot (red) be reached by mutating cytosines and guanines to any other base for each of the 64 codons, highlighting that only 32% of missense mutations are achievable with targeted cytidine deaminases. The white area depicts the original amino acid identity.



Extended Data Fig. 2 | The direction of EvolvR-mediated mutagenesis relative to the gRNA is dependent on which strand is nicked. Our previous fluctuation analysis in Fig. 1e demonstrated that nCas9(D10A)-PolI3M mutates a window 3' of the nick site. Here we directly tested whether mutations are generated 5' of the nick site using a different gRNA. Because DNA polymerases synthesize in the 5'-to-3' direction, we anticipated that nCas9(D10A)-PolI3M would not provide an elevated mutation rate 5' of the nick site. We indeed found that expressing a guide RNA which targeted nCas9(D10A)-PolI3M to nick 16 nucleotides 3' from the nonsense mutation (indicated by a red cross) did not show targeted mutagenesis. We hypothesized that we could induce targeted mutagenesis using the same gRNA by using a Cas9 variant harbouring the H840A mutation, which nicks the DNA strand non-complementary to the gRNA, rather than the D10A mutation, which nicks the strand complementary to the gRNA. nCas9(H840A)-PolI3M increased the mutation rate 16 nucleotides 3' from the nick by 52-fold compared to the global mutation rate of cells expressing an off-target gRNA. We used the D10A nCas9 variant for all subsequent experiments. Data are mean \pm 95% confidence intervals from ten biologically independent samples. * $P < 0.0001$; two-sided Student's t -test.

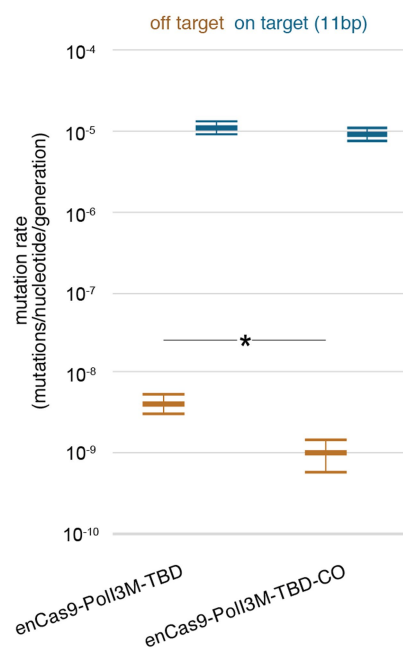


Extended Data Fig. 3 | PolI5M elevates mutation rates 1 nucleotide, but not 11 nucleotides, from the nick compared to PolI3M. PolI3M with additional F742Y and P796H mutations (PolI5M) elevates the mutation rate 33-fold 1 nucleotide from the nick compared to PolI3M. PolI5M did not have a higher mutation rate than PolI3M 11 nucleotides from the nick. Data are mean \pm 95% confidence intervals from ten biologically independent samples. * $P < 0.0001$; two-sided Student's t -test.

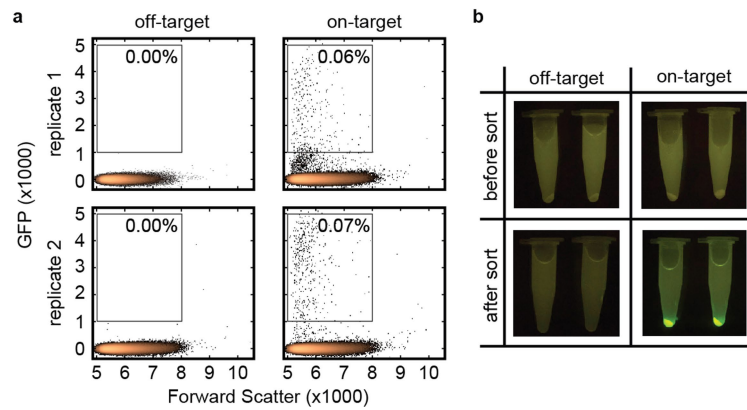


Extended Data Fig. 4 | Fusing a highly processive DNA polymerase to enCas9 increases the target window length. PolI was exchanged for a more processive and higher-fidelity bacteriophage Phi29 DNA polymerase (Phi29). Owing to Phi29 not having a flap endonuclease, residues 1–325 of PolI were inserted between enCas9 and Phi29. Using gRNAs targeting different distances from the nonsense mutation, we found that Phi29 with two previously reported fidelity-reducing mutations (N62D and L384R) elevated the mutation rate 56 nucleotides from the nick compared to the global mutation rate^{28,29}. When we expressed Phi29's single-stranded binding protein (ssb), which is known to improve the activity of Phi29, we observed an elevation in the targeted mutation rate³⁰. Finally, because

the activity of Phi29 is known to decrease at temperatures above 30 °C and the fluctuation analysis was performed at 37 °C, we added mutations previously reported to improve the thermostability of Phi29 (iPhi29) and observed a targeted mutation rate 347 nucleotides from the nick site that was significantly greater than the global mutation rate³¹. Unfortunately, mutations decreasing Phi29's fidelity are known to decrease its processivity explaining our inability to identify Phi29 variants that retain high processivity while offering as high of a mutation rate as PolI3M²⁸. Data are mean \pm 95% confidence intervals from ten biologically independent samples. * $P < 0.0001$; two-sided Student's t -test.

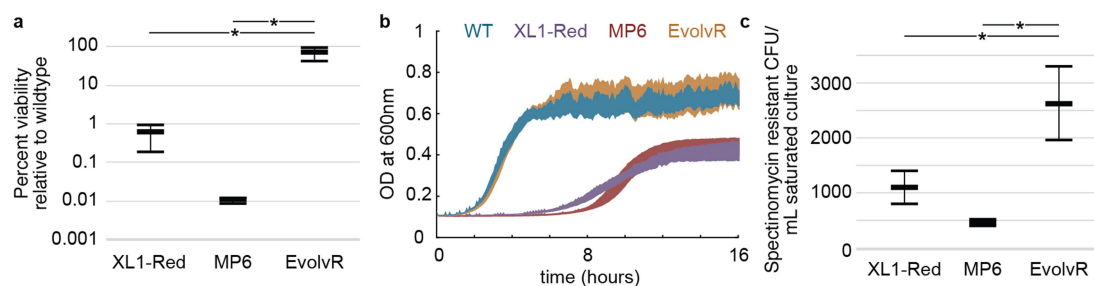


Extended Data Fig. 5 | Removing internal ribosome binding sequences decreases EvolvR-mediated off-target mutagenesis. enCas9-PolI3M-TBD was codon optimized to remove strong ribosome binding sites in the EvolvR coding sequence that were predicted to produce an untethered DNA polymerase. The off-target mutation rate decreased 4.14-fold when expressing enCas9-PolI3M-TBD-CO compared to enCas9-PolI3M-TBD ($P = 0.000482$) whereas the on-target mutation rate only decreased 1.23-fold. Data are mean \pm 95% confidence intervals from ten biologically independent samples. * $P < 0.0001$; two-sided student's t -test.



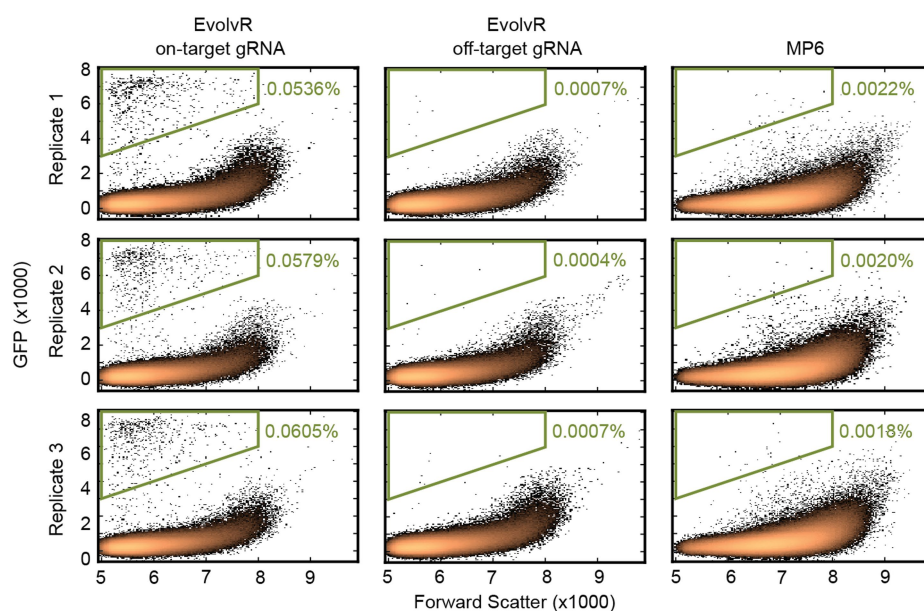
Extended Data Fig. 6 | EvolvR-mediated mutagenesis can be coupled with a non-selectable genetic screen. **a**, To test the capability for coupling EvolvR-mediated mutagenesis with a non-selectable genetic screen, we designed a target plasmid containing a GFP cassette with an early termination codon in the GFP coding sequence (pTarget-GFP*). After co-transforming pEvolvR with pTarget-GFP* and growing for 24 h, we analysed and sorted the GFP-positive fraction. In the two replicates expressing an off-target gRNA, we did not detect or sort any GFP cells. By contrast, for the two replicates expressing a gRNA nicking four nucleotides

away from the chain-terminating mutation in the coding sequence of GFP, we found that 0.06% and 0.07% of the total cells were GFP positive. These results agree with sequencing outcomes from Fig. 1b, which showed that expressing nCas9–PolI3M for 24 h produces substitutions in the target region at frequencies between 0.5% to 1%. **b**, After culturing the sorted populations, both replicates expressing an off-target gRNA did not show growth, whereas both replicates expressing the on-target gRNA grew bright green.



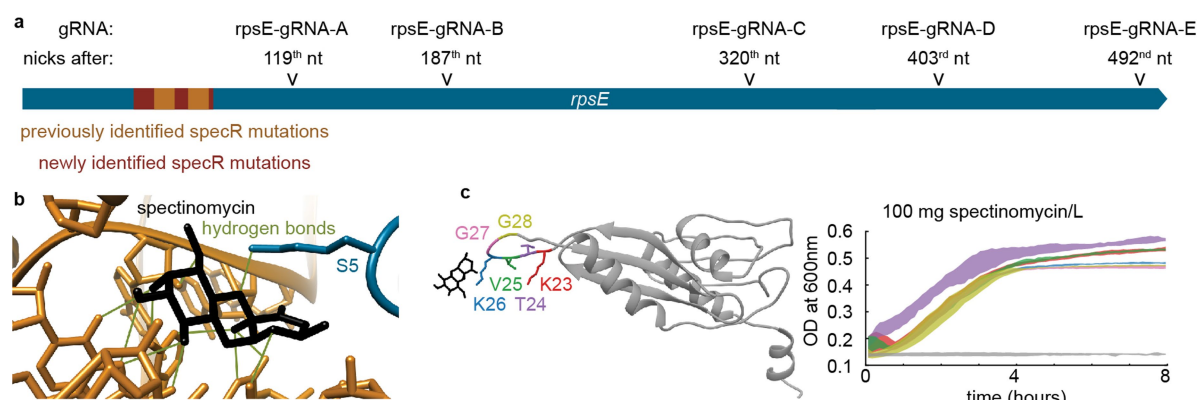
Extended Data Fig. 7 | EvolvR enables targeted genome diversification without affecting viability or growth rate. **a**, The viability of TG1 *E. coli* expressing EvolvR targeted to the essential *rpsE* gene was significantly higher than TG1 *E. coli* transformed with the MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose (a previously developed plasmid for continuous non-targeted mutagenesis³², $P = 0.0108$) as well as XL1-Red *E. coli* (a previously developed strain for continuous non-targeted mutagenesis³³, $P = 0.0105$). Viability was measured relative to TG1 *E. coli* transformed with an empty control plasmid. Data are mean \pm s.d. from three biologically independent samples. $*P < 0.05$; two-tailed t -test. **b**, TG1 *E. coli* transformed with an empty control

plasmid and TG1 *E. coli* transformed with pEvolvR targeting the *rpsE* gene resulted in similar growth curves whereas XL1-Red *E. coli* and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose grew much slower and saturated at lower final optical densities. Shaded area represents mean \pm s.d. from three biologically independent samples. **c**, The spectinomycin-resistant CFUs per ml saturated culture of TG1 *E. coli* targeting EvolvR to the *rpsE* gene was significantly higher than XL1-Red *E. coli* ($P = 0.022$) and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose ($P = 0.0049$). Data are mean \pm s.d. from three biologically independent samples. $*P < 0.05$; two-tailed t -test.



Extended Data Fig. 8 | EvolvR-mediated mutagenesis performs better than a previous non-targeted diversification technique. To compare the performance of EvolvR and the previously developed non-targeted mutagenesis plasmid MP6 in screen-based directed evolution applications, we co-transformed pEvolvR (enCas9–PolI3M–TBD) or MP6 with a target plasmid containing a GFP cassette with an early termination codon in the

GFP coding sequence (pTarget-GFP*). The cultures expressing EvolvR were grown for 24 h and the MP6 cultures followed a two day growth-induction protocol as previously described. Flow cytometry revealed that cultures expressing EvolvR and an on-target gRNA resulted in 28-fold more GFP-positive cells than MP6 cultures.



Extended Data Fig. 9 | Locations of gRNA targets relative to the *rpsE* gene and mutations in ribosomal protein S5 that confer spectinomycin resistance. **a**, enCas9–PolI3M–TBD was targeted to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair of the 504-bp *rpsE* coding sequence. The locations of the previously identified *rpsE* mutations that provide spectinomycin resistance are coloured orange, and the region where we identified new spectinomycin-resistance mutations is highlighted in red. **b**, The mutations that we discovered confer spectinomycin resistance would be expected to move Lys26 (which is predicted to hydrogen bond with spectinomycin) relative to the spectinomycin-binding

pocket. We hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket remove that hydrogen bond and destabilize the interaction of spectinomycin with the ribosome, thereby conferring spectinomycin resistance. **c**, Therefore, we tested whether deleting any single amino acid between residues 16 and 35 confers spectinomycin resistance. We found that deleting residues 23, 24, 25, 26, 27 or 28 provides spectinomycin resistance whereas deleting any of the residues between 16 and 22 or 29 and 35 does not. These results support the hypothesis that one mechanism of resistance to spectinomycin is disruption of the interaction between Lys26 and spectinomycin. Data are mean \pm s.d. from three biologically independent samples.

Extended Data Table 1 | Comparison of *E. coli* diversification methods

<i>E. coli</i> Diversification Method	Host/culture requirements	Targetability	Ease of use	<i>In vivo</i> /Continuous?	When to use
XL1-Red ³³	Specific Strain	None	Transformation	Yes	Continuous whole genome evolution; target is unknown
MP6 ³²	None	None	Transformation	Yes	Continuous whole genome evolution of any strain; target is unknown
Orthogonal polymerase/plasmid ³	Specific strain and culture requirements	Plasmid	Transformation	Yes	Continuous plasmid evolution; target must be located next to the origin of replication of a specific plasmid
PACE ²	Specific strain and culture requirements	Phage genome	Custom turbidostat operation, bacteriophage propagation	Yes	Continuous engineered phage genome evolution; target must be inserted within phage genome, and target activity must be coupled to phage propagation
MAGE ⁶	Specific strain	User-defined targets	Recombination machinery induction and high-efficiency transformation of oligonucleotide library	No	Generating rationally designed, discrete, user-defined libraries of recombineering strains
EvolvR	None	User-defined targets	Transformation	Yes	Continuous diversification of user-defined genomic loci in any strain

Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth

Jonathan B. Baell^{1,2*}, David J. Leaver¹, Stefan J. Hermans³, Gemma L. Kelly^{4,5}, Margs S. Brennan^{4,5}, Natalie L. Downer⁴, Nghi Nguyen¹, Johannes Wichmann^{4,5}, Helen M. McRae^{4,5}, Yuqing Yang^{4,5}, Ben Cleary¹, H. Rachel Lagiakos^{1,6}, Stephen Mieruszynski^{4,5}, Guido Pacini⁴, Hannah K. Vanyai^{4,5}, Maria I. Bergamasco^{4,5}, Rose E. May⁴, Bethany K. Davey^{4,6}, Kimberly J. Morgan^{4,5}, Andrew J. Sealey^{4,5}, Beinan Wang^{4,5,7}, Natasha Zamudio^{4,5}, Stephen Wilcox^{4,5}, Alexandra L. Garnham^{4,5}, Bilal N. Sheikh^{4,5}, Brandon J. Aubrey^{4,5}, Karen Doggett^{4,5}, Matthew C. Chung³, Melanie de Silva^{4,6}, John Bentley⁸, Pat Pilling⁸, Meghan Hattarki⁸, Olan Dolezal⁸, Matthew L. Dennis⁸, Hendrik Falk^{4,5,6}, Bin Ren⁸, Susan A. Charman⁹, Karen L. White⁹, Jai Rautela^{4,5}, Andrea Newbold¹⁰, Edwin D. Hawkins^{4,5}, Ricky W. Johnstone¹⁰, Nicholas D. Huntington^{4,5}, Thomas S. Peat⁸, Joan K. Heath^{4,5}, Andreas Strasser^{4,5}, Michael W. Parker^{3,11}, Gordon K. Smyth^{4,12}, Ian P. Street^{4,5,6}, Brendon J. Monahan^{4,5,6}, Anne K. Voss^{4,5,13*} & Tim Thomas^{4,5,13*}

Acetylation of histones by lysine acetyltransferases (KATs) is essential for chromatin organization and function¹. Among the genes coding for the MYST family of KATs (KAT5–KAT8) are the oncogenes KAT6A (also known as MOZ) and KAT6B (also known as MORF and QKF)^{2,3}. KAT6A has essential roles in normal haematopoietic stem cells^{4–6} and is the target of recurrent chromosomal translocations, causing acute myeloid leukaemia^{7,8}. Similarly, chromosomal translocations in KAT6B have been identified in diverse cancers⁸. KAT6A suppresses cellular senescence through the regulation of suppressors of the *CDKN2A* locus^{9,10}, a function that requires its KAT activity¹⁰. Loss of one allele of KAT6A extends the median survival of mice with MYC-induced lymphoma from 105 to 413 days¹¹. These findings suggest that inhibition of KAT6A and KAT6B may provide a therapeutic benefit in cancer. Here we present highly potent, selective inhibitors of KAT6A and KAT6B, denoted WM-8014 and WM-1119. Biochemical and structural studies demonstrate that these compounds are reversible competitors of acetyl coenzyme A and inhibit MYST-catalysed histone acetylation. WM-8014 and WM-1119 induce cell cycle exit and cellular senescence without causing DNA damage. Senescence is INK4A/ARF-dependent and is accompanied by changes in gene expression that are typical of loss of KAT6A function. WM-8014 potentiates oncogene-induced senescence in vitro and in a zebrafish model of hepatocellular carcinoma. WM-1119, which has increased bioavailability, arrests the progression of lymphoma in mice. We anticipate that this class of inhibitors will help to accelerate the development of therapeutics that target gene transcription regulated by histone acetylation.

In a screen of 243,000 diverse small-molecule compounds¹², we obtained the acylsulfonhydrazide compound CTx-0124143, a competitive KAT6A inhibitor (half-maximal inhibitory concentration (IC₅₀) 0.49 μ M) in biochemical assays¹². Medicinal chemistry optimization yielded the compound WM-8014 with an IC₅₀ value of 8 nM (Fig. 1a, Supplementary Table 1), representing a 60-fold increase in inhibitory activity towards KAT6A. This was consistent with the binding affinity measured by surface plasmon resonance (SPR; equilibrium dissociation constant (*K*_D) 5 nM; Fig. 1a, Extended Data Fig. 1). WM-8014 inhibits predominantly the closely related proteins KAT6A

and KAT6B (IC₅₀ 8 nM and 28 nM, respectively), and is more than tenfold less active against KAT7 and KAT5 (IC₅₀ 342 nM and 224 nM, respectively; Fig. 1b, Supplementary Table 1). Kinetic binding curves obtained from SPR demonstrated that the interaction of this class of compounds with immobilized proteins was fully reversible and consistent with a single-site binding interaction. The interaction of WM-8014 with KAT6A and KAT7, although relatively strong, was in both cases driven by fast association kinetics (association rate constant (*k*_a) > 1 × 10⁶ M⁻¹ s⁻¹), whereas the dissociation kinetics (dissociation rate constant (*k*_d) ~ 4 × 10⁻² for KAT6A and 17 × 10⁻² s⁻¹ for KAT7) were indicative of a relatively short lifespan of the binary complex (Extended Data Fig. 1). WM-8014 displayed an order of magnitude weaker binding to KAT7 than to KAT6A (*K*_D 52 nM versus 5.1 nM, respectively) (Fig. 1b, Extended Data Fig. 1). We also generated an inactive analogue, WM-2474 (Fig. 1a, Supplementary Table 1). Notably, these compounds were almost inactive against KAT8, and no inhibition was observed for the more distantly related lysine acetyltransferases KAT2A, KAT2B, KAT3A and KAT3B (Fig. 1b, c, Supplementary Table 1).

WM-8014 has desirable, drug-like physicochemical properties (Supplementary Table 2). It is completely stable in cell culture medium (10% fetal calf serum); however, relatively high protein binding (97.5%) in this medium reduces its free concentration. Although WM-8014 has relatively low solubility in water (8–16 μ M), it could readily permeate Caco-2 cells (apparent permeability coefficient (*P*_{app}) 78 ± 13 × 10⁻⁶ cm s⁻¹). Testing of WM-8014 at 1 μ M and 10 μ M revealed no notable affinity for a pharmacological panel of 158 diverse biological targets; only eight targets were affected by more than 50% (Supplementary Table 3).

We solved the crystal structures of a modified MYST histone acetyltransferase domain (MYST^{Cryst}) in complex with WM-8014 (1.85 Å resolution, Fig. 1d–f, Extended Data Fig. 1, Supplementary Table 4) or acetyl coenzyme A (acetyl-CoA; 1.95 Å resolution, Fig. 1g). The WM-8014 molecule occupies the acetyl-CoA-binding site on MYST^{Cryst}, being partially enclosed between the α -helix formed by residues D685 to R704 and the loop extending from Q654 to G657. The MYST^{Cryst}–acetyl-CoA complex adopts a globular fold (Fig. 1g), as seen in previously reported structures¹³, with a root mean square deviation

¹Medicinal Chemistry Theme, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia. ²School of Pharmaceutical Sciences, Nanjing Tech University, Nanjing, China. ³ACRF Rational Drug Discovery Centre, St Vincent's Institute of Medical Research, Fitzroy, Victoria, Australia. ⁴The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria, Australia. ⁵Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia. ⁶Cancer Therapeutics CRC, Parkville, Victoria, Australia. ⁷School of Pharmaceutical Sciences, Tsinghua University, Beijing, China. ⁸Commonwealth Scientific and Industrial Research Organisation (CSIRO), Biomedical Program, Parkville, Victoria, Australia. ⁹Centre for Drug Candidate Optimisation, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia. ¹⁰The Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ¹¹Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia. ¹²Department of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia. ¹³These authors jointly supervised this work: Anne K. Voss, Tim Thomas. *e-mail: Jonathan.Baell@monash.edu; avoss@wehi.edu.au; tthomas@wehi.edu.au

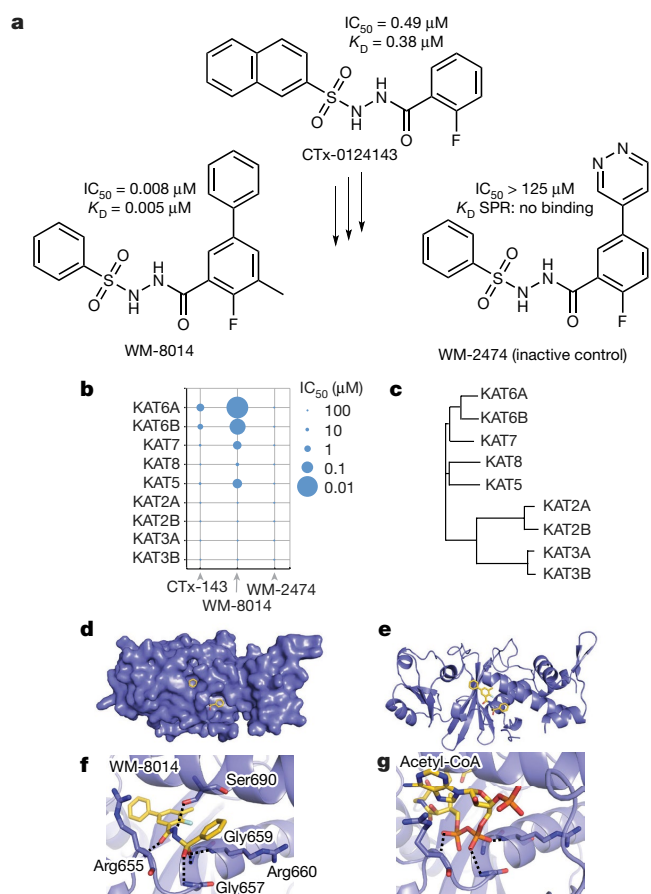


Fig. 1 | Development of an inhibitor of the MYST family of lysine acetyltransferases. a, Schematic summary of the medicinal chemistry optimization of high-throughput screening hit CTx-0124143, which resulted in WM-8014 and the inactive compound WM-2474. The IC₅₀ values (determined by biochemical assays) and equilibrium dissociation constants (K_D, determined by SPR) are shown for KAT6A. **b**, Histone acetyltransferase inhibition assay (competition of compound with acetyl-CoA) of CTx-0124143, WM-8014 and WM-2474. The areas of the circles reflect the IC₅₀ values as indicated, assayed at the Michaelis constant (K_m) of acetyl-CoA for each KAT tested. **c**, Dendrogram showing the relationship between major KAT families based on sequence differences in the acetyltransferase domain. **d–g**, Crystal structures of WM-8014 and acetyl-CoA bound to the MYST lysine acetyltransferase domain (MYST^{Cryst}; see Extended Data Fig. 1). PDB codes: 6BA2 and 6BA4, respectively. **d**, Space-filling model showing WM-8014 in the acetyl-CoA-binding pocket of MYST^{Cryst}. **e**, Ribbon diagram of MYST^{Cryst} (blue) showing WM-8014 (yellow, with element colouring) bound to the acetyl-CoA-binding site. **f**, Ribbon diagram of MYST^{Cryst} showing key amino acids interacting with WM-8014 (yellow, with element colouring). Hydrogen bonds are shown as dashed lines. **g**, Ribbon diagram showing acetyl-CoA (yellow, with element colouring) bound to MYST^{Cryst}. Means of two experiments are shown for the IC₅₀ values in **a** and **b**. SPR experiments in **a** were repeated four times.

(r.m.s.d.) of 0.6 Å, and is nearly identical to the MYST^{Cryst}–WM-8014 complex (r.m.s.d. of 0.3 Å for all aligned atoms). Accordingly, the core acylsulfonylhydrazide moiety of WM-8014 makes similar hydrogen bonds to MYST^{Cryst} as does the diphosphate group of acetyl-CoA (Fig. 1f, g). This includes hydrogen bonds to the main-chain atoms of R655, G657 and R660—identical to acetyl-CoA—as well as additional hydrogen bonds to G659 and S690 (Extended Data Fig. 1). The biphenyl group of WM-8014 extends further into the acetyl-CoA-binding pocket, which enables van der Waals interactions with residues L601, I647, I649, S684 and L686 of MYST^{Cryst} (Extended Data Fig. 1). WM-8014 therefore competes directly with acetyl-CoA in the substrate-binding domain.

Because KAT6A suppresses senescence^{9,10}, we examined the ability of WM-8014 to induce cell cycle arrest in embryonic day (E)14.5 mouse embryonic fibroblasts (MEFs). Cells treated with WM-8014 failed to proliferate after 10 days of treatment (Fig. 2a; IC₅₀ 2.4 μM), with similar kinetics to Cre-recombinase *Kat6a* recombination (Fig. 2b). Higher doses of WM-8014 (up to 40 μM) did not accelerate growth arrest, which after 8 days of treatment was irreversible (Extended Data Fig. 2). The inactive compound WM-2474 did not affect cell proliferation. Cell cycle analysis showed an increase in the proportion of cells in G0/G1 after 6 days of treatment and a corresponding reduction in cells in G2/M and S phases, both in Fucci cells¹⁴ and in 5-bromo-2'-deoxyuridine (BrdU) incorporation assays (Fig. 2c, Extended Data Fig. 2).

RNA sequencing (RNA-seq) of MEFs treated with WM-8014 revealed a signature of cellular senescence, including upregulated expression of *Cdkn2a* mRNA and decreased expression of *Cdc6*, which is a KAT6A target gene⁹ and a regulator of DNA replication¹⁵ (Fig. 2d; day 10: false discovery rate (FDR) < 10^{−6}). A substantial increase in β-galactosidase activity—a marker of senescent cells—was also observed (Fig. 2e), accompanied by morphological changes typical of senescence (Extended Data Fig. 2). WM-8014 caused a concentration-dependent reduction in the level of *E2f2* mRNA (adjusted (adj.) R² = 0.73; P < 0.0005) and *Cdc6* mRNA (adj. R² = 0.5; P = 0.002), accompanied by upregulation of both splice products of the *Cdkn2a* locus, *Ink4a* and *Arf* (day 10: P < 0.0005 and P = 0.005, respectively; Extended Data Fig. 3). Notably, MEFs treated for 4 days or 10 days with 10 μM WM-8014, the control compound WM-2474 or DMSO vehicle control showed no change in the levels of γH2A.X (Extended Data Fig. 4), which suggests that cell cycle arrest was not a consequence of DNA damage. No increase in apoptosis or necrosis was seen (Extended Data Fig. 4). Treatment of either *Trp53*-null MEFs (*Trp53*^{−/−}) or *Cdkn2a*-null (*Ink4a*^{−/−} *Arf*^{−/−}) MEFs with WM-8014 had a minor effect and no effect on cell proliferation, respectively (Fig. 2f, Extended Data Fig. 2). These results show that WM-8014 acts through the p16^{INK4A}–p19^{ARF} pathway, causing irreversible cell cycle exit leading to senescence, and does not have a general cytotoxic effect.

KAT7 is essential for global histone 3 lysine 14 (H3K14) acetylation¹⁶. By contrast, KAT6A regulates H3K9 acetylation only at target loci^{17,18}. We determined the effects of WM-8014 on global levels of acetylation at H3K9 and H3K14 by western blot after 5 days of treatment. Treatment with 10 μM WM-8014 caused a 49% decrease in the global levels of H3K14ac but, as expected on the basis of the locus-specific roles of KAT6A^{17,18}, did not significantly affect the global levels of H3K9ac (Fig. 3a, b; all gel source data in Supplementary Fig. 1). The effects of WM-8014 on global H3K14ac levels were concentration-dependent (Fig. 3b; H3K14ac/H4 ratio regressed on log concentration of WM-8014; adj. R² = 0.76, P < 0.001; IC₅₀ 1.2 μM). RNA-seq showed a strong correlation between the changes in gene expression seen in *Kat6a*^{−/−} MEFs compared with *Kat6a*^{+/+} MEFs and the genes differentially expressed after WM-8014 treatment (WM-8014 compared with inactive WM-2474), with a 2.6-fold enrichment in upregulated genes (FDR = 0.0001; Fig. 3c) and a 2.1-fold enrichment in downregulated genes (FDR = 0.0001; Fig. 3c), and gene expression signatures characteristic of cellular senescence (Extended Data Fig. 5). Loss of KAT6A results in the downregulation of *E2f2*, *Ezh2* and *Melk*⁹. Similarly, treatment with WM-8014 caused significant downregulation of *Ezh2*, *Melk* and *E2f2* mRNA levels compared with controls (Fig. 3d), as determined by RNA-seq (Extended Data Fig. 5) and confirmed by quantitative reverse-transcription PCR (RT-qPCR) (Extended Data Fig. 3). After treatment with WM-8014, there was a reduction of H3K9ac at the transcription start sites of these genes (Fig. 3e). Therefore, the treatment of cells with high concentrations of WM-8014 directly inhibits global H3K14 acetylation catalysed by KAT7, as well as KAT6A-specific H3K9 acetylation at transcription start sites.

Because WM-8014 induced cellular senescence, we reasoned that it might exacerbate oncogenic RAS-induced senescence. Accordingly, MEFs that express HRAS^{G12V}, a constitutively active form of RAS, were more sensitive to the induction of cell cycle arrest by WM-8014

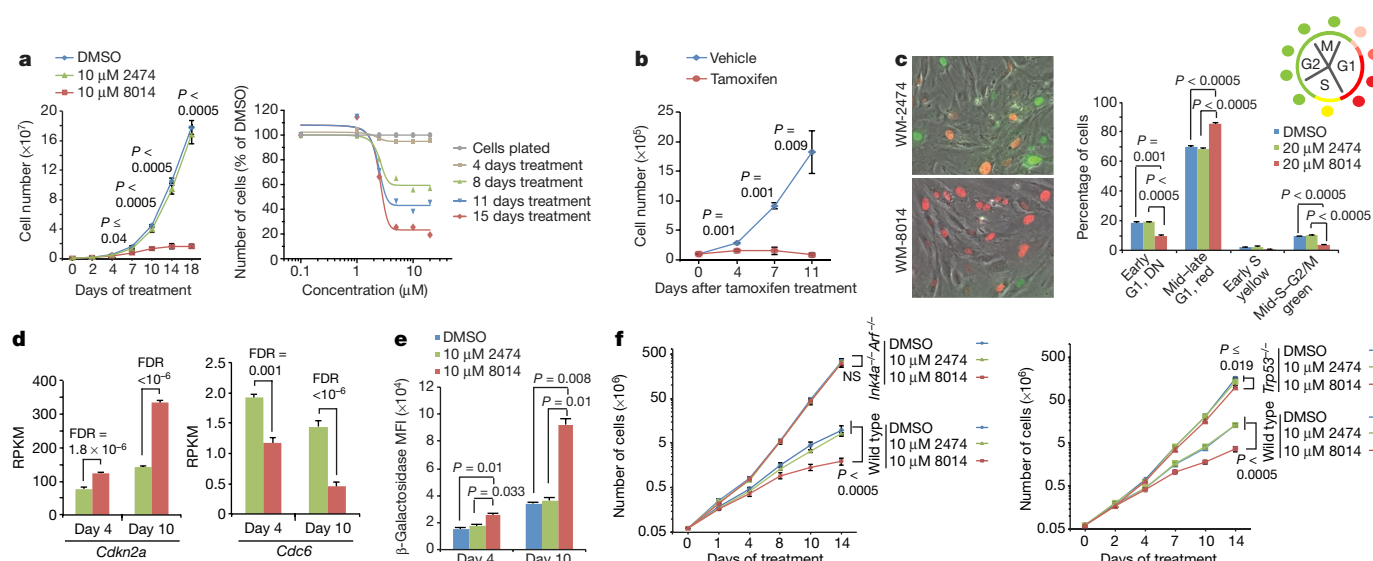


Fig. 2 | Treatment of MEFs with WM-8014 leads to cellular senescence.

a, Left, effects of WM-8014 compared with the inactive compound WM-2474 or DMSO vehicle control on cell growth of MEFs grown in 3% O₂. Right, effects of the dose of WM-8014 and the duration of treatment. **b**, Effects of acute genetic deletion of *Kat6a* on the growth of MEFs. Loss of KAT6A function was induced by nuclear translocation of Cre-recombinase using tamoxifen on MEFs isolated from *Kat6a^{lox/lox} Rosa^{CreERT2}* and control *Rosa^{CreERT2}* embryos. **c**, Left, epifluorescence phase-contrast images of Fucci MEFs after 6 days of treatment with 20 μM WM-2474 (top) and 20 μM WM-8014 (bottom). Right, the percentage of Fucci MEFs in each stage of the cell cycle after 6 days of treatment with 10 μM WM-8014, 10 μM WM-2474 or DMSO vehicle control, as quantified by flow-cytometry analysis. DN, double negative. **d**, mRNA levels of *Cdkn2a* (coding for cell cycle regulators p16^{INK4A} and p19^{ARF}) (left) and the KAT6A target gene *Cdc6* (right) in MEFs treated for 4 days

and 10 days with 10 μM WM-8014 or 10 μM WM-2474 control, assessed by RNA-seq. RPKM, reads per kilobase per million reads. **e**, Flow-cytometry assessment (mean ± s.e.m. of median fluorescence intensity (MFI)) of senescence-associated β-galactosidase activity in MEFs after 4 and 10 days of treatment with 10 μM WM-8014, 10 μM WM-2474 or DMSO vehicle control. **f**, Growth of MEFs lacking p16^{INK4A} and p19^{ARF} (left) and of MEFs lacking p53 (right) compared with wild type after treatment with WM-8014, DMSO vehicle control or WM-2474. *n* = 3 independent MEF isolates per treatment group and genotype. Data are mean ± s.e.m. Data were analysed by one-way ANOVA followed by Bonferroni post hoc test (**a** (left), **b**, **c**), non-linear regression curve fit (**a**, right) or two-way ANOVA (**f**) with treatment and with or without treatment duration as the independent factors. RNA-seq data (**d**) were analysed as described in the Supplementary Methods.

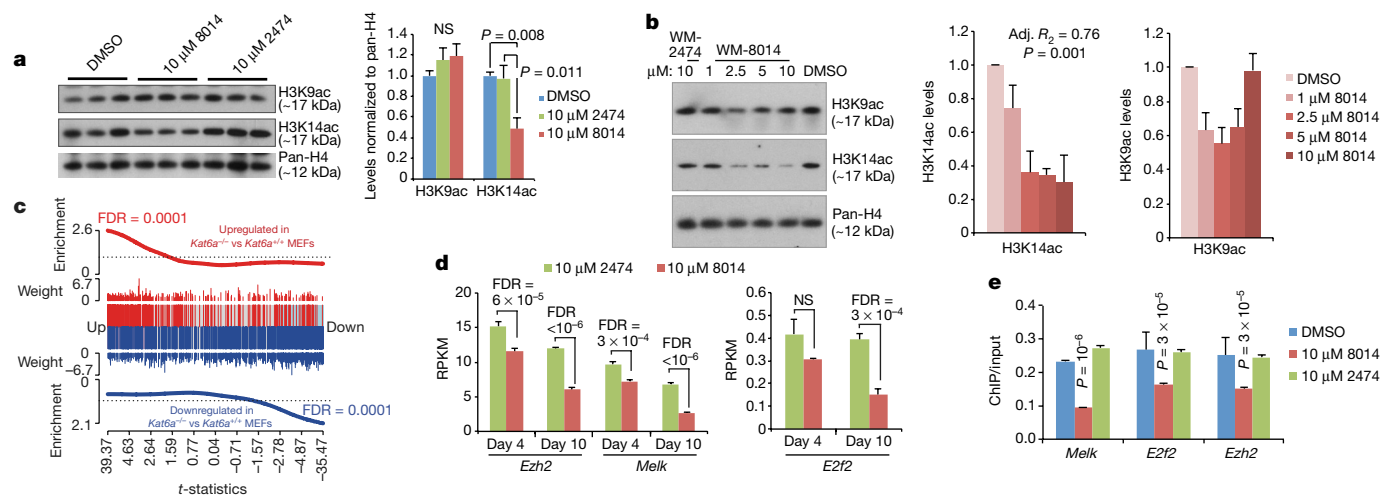


Fig. 3 | Treatment of cells with WM-8014 leads to a reduction in acetylation of specific histone lysine residues and changes in gene expression that resemble the genetic loss of KAT6A. **a**, Western blot detection of H3K14ac or H3K9ac in MEFs treated with 10 μM WM-8014, 10 μM WM-2474 or DMSO for 5 days. The densitometric analysis is presented on the right. *n* = 6 (H3K14ac) and *n* = 9 (H3K9ac) independent cultures per treatment group. **b**, Western blot of MEFs treated with increasing doses of WM-8014 and controls as indicated. Densitometric analysis is presented on the right. *n* = 3 independent experiments. Histone acetylation levels were regressed on the log₁₀ of the WM-8014 concentration. H3K14ac and H3K9ac levels were normalized to pan-H4 levels and DMSO treatment. **c**, Barcode plot in which genes that are differentially up- or downregulated in *Kat6a^{-/-}* versus *Kat6a^{+/+}* MEFs (that is, after genetic deletion of KAT6A) are compared with genes differentially expressed in MEFs treated with WM-8014 versus WM-2474. Combined results of day 4 and day 10 treatment, ROAST *P* = 0.0001;

MEF isolates from individual E12.5 embryos, namely from *n* = 3 *Kat6a^{-/-}* and 2 *Kat6a^{+/+}* embryos, as well as 3 MEF isolates from 3 wild-type embryos treated with either WM-8014 or WM-2474. **d**, *Ezh2*, *Melk* and *E2f2* mRNA levels measured by RNA-seq in MEFs treated for 4 days and 10 days with 10 μM WM-8014 or 10 μM control WM-2474 (*n* = 3 MEF isolates from 3 wild-type embryos treated with either WM-8014 or WM-2474). **e**, Anti-H3K9ac chromatin immunoprecipitation followed by qPCR detection of transcription start sites of genes after treatment with DMSO 10 μM WM-8014 or WM-2474 for 3 days. The results of one of four experiments are shown; total *n* = 16 cultures per treatment group in 4 experiments. Data are mean ± s.e.m. (with the exception of **e**, mean ± s.d.) and were analysed by one-way ANOVA followed by Bonferroni post hoc test (**a**), by regression analysis (**b**) or by *t*-test comparing WM-8014 to WM-02474 (**e**). The RNA-seq analysis (**c**, **d**) is described in the Supplementary Methods.

(Extended Data Fig. 6). We then examined the effects of WM-8014 in a zebrafish model¹⁹ of KRAS^{G12V}-driven hepatocellular overproliferation. We observed a significant, concentration-dependent reduction in liver volume in response to treatment with WM-8014, and a substantial reduction in hepatocytes in S phase (Extended Data Fig. 6). Notably, WM-8014 did not impair the growth of the normal liver, demonstrating that the inhibitory effects of WM-8014 were specific to hepatocytes that express oncogenic RAS. Treatment with WM-8014 was found to robustly upregulate the cell cycle regulators *Cdkn2a* and *Cdkn1a* in hepatocytes that express oncogenic KRAS^{G12V}, but not control hepatocytes. Therefore, WM-8014 potentiates oncogene-induced senescence, but it does not affect normal hepatocyte growth.

The progression of lymphoma is highly dependent on KAT6A, as *Kat6a* heterozygous mice are protected from early-onset MYC-driven lymphoma¹¹. However, the high levels of plasma-protein binding exhibited by WM-8014 (Supplementary Table 2) precluded in vivo studies in mice. Development of derivatives of WM-8014 resulted in WM-1119, which has reduced plasma-protein binding (Fig. 4a; Supplementary Table 2). The interaction of WM-1119 with KAT6A is similar to that of WM-8014: it is characterized by strong reversible binding (K_D 2 nM, compared with 5 nM for WM-8014; Extended Data Fig. 7) that is competitive with acetyl-CoA, and driven by fast association kinetics ($k_a > 1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$; Extended Data Fig. 7). The structure of MYST^{Cryst} in complex with WM-1119 was solved (Extended Data Fig. 7, Supplementary Table 5) and was found to be almost identical to that of MYST^{Cryst}-WM-8014, with an r.m.s.d. for aligned main-chain atoms of 0.2 Å. There are two key differences between the complexes: an additional hydrogen bond is formed between the WM-1119 pyridine nitrogen and the main chain at I649 that is not present in the complex with WM-8014 (Extended Data Fig. 7), and the hydrophobic interaction that exists between the meta-methyl of the biphenyl group of WM-8014 and I663 is not present in the complex with WM-1119. WM-1119 is 1,100-fold and 250-fold more active against KAT6A than against KAT5 or KAT7, respectively (Fig. 4a, Extended Data Fig. 7), and so shows greater specificity for KAT6A than does WM-8014. The testing of WM-1119 at 1 μM and 10 μM against a pharmacological panel of 159 diverse biological targets revealed no affinity (Supplementary Table 6). Treatment of MEFs with WM-1119 resulted in cell cycle arrest in G1 and a senescence phenotype similar to that seen upon treatment with WM-8014 (Extended Data Fig. 8). Notably, the activity of WM-1119 in this cell-based assay is an order of magnitude greater than WM-8014 and WM-1119 is able to induce cell cycle arrest at 1 μM .

To test inhibitors of KAT6A in a cancer model, we investigated the effect of WM-1119 and WM-8014 on the proliferation of lymphoma cells. We selected the B cell lymphoma cell line EMRK1184, which was isolated from mice with a tumour resulting from the expression of *Myc* under the control of the IgH enhancer²⁰, because it expressed the *Cdkn2a*-locus-encoded ARF and wild-type p53 (Extended Data Fig. 9). Treatment with WM-8014 or WM-1119 inhibited the proliferation of the EMRK1184 lymphoma cells in vitro (Fig. 4b); RNA-seq and western blot analysis showed that treatment with WM-1119 resulted in increased levels of *Cdkn2a* and *Cdkn2b* mRNA and p16^{INK4a} and p19^{ARF} proteins, as well as a delayed increase in *Cdkn1a* mRNA (Extended Data Fig. 9). WM-1119 (IC_{50} 0.25 μM) was ninefold more active than WM-8014 (IC_{50} 2.3 μM ; Fig. 4b), as expected on the basis of reduced protein binding (Supplementary Table 2).

We tested the effectiveness of KAT6 inhibitors in the treatment of lymphoma in mice. Male C57BL/6-albino (B6(Cg)-Tyr^{c-2J}/J) mice were injected intravenously with 100,000 EMRK1184 cells transfected with a luciferase-expression construct. Lymphoma growth was monitored using the IVIS imaging system. Three days after the lymphoma-cell transplant, all mice showed luciferase activity (Fig. 4c), which indicated the expansion of lymphoma cells. Mice were then divided randomly into WM-1119-treatment and vehicle-control groups. Because WM-1119 is rapidly cleared after intraperitoneal injection, with the plasma concentration decreasing to below 1 μM after 4–6 h

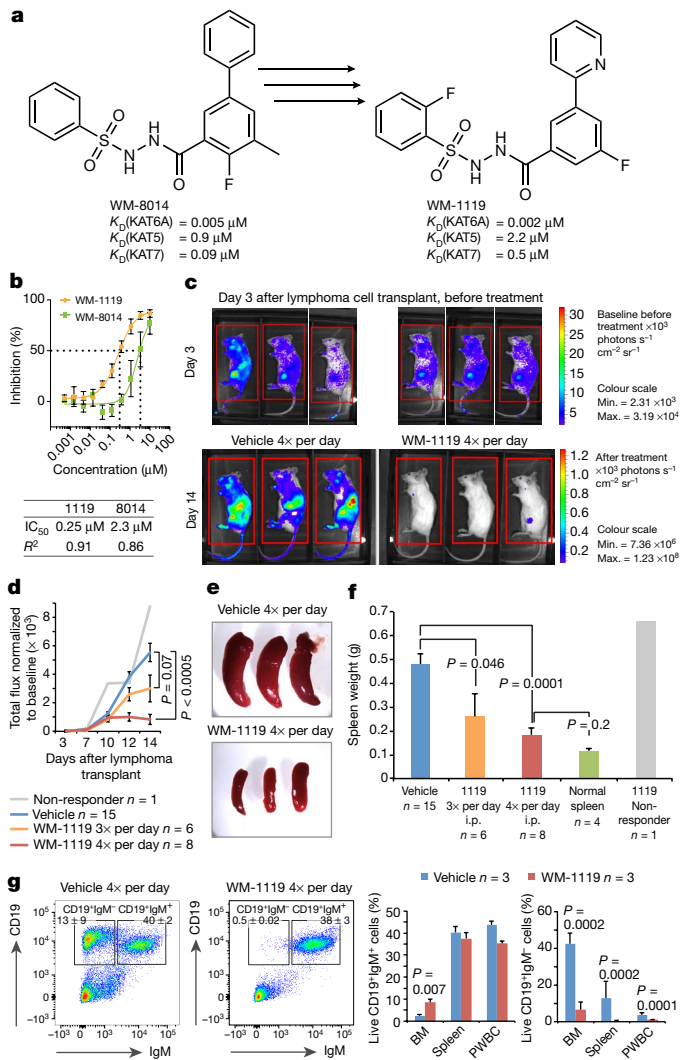


Fig. 4 | Treatment with WM-1119 arrests lymphoma growth. **a**, Medicinal chemistry optimization of WM-8014 resulted in compound WM-1119. The binding data (obtained by SPR) for the interaction of WM-1119 with immobilized KAT6A, KAT7 and KAT5 are compared with the interaction data for WM-8014. **b**, Growth inhibition assays of *Eμ-Myc* lymphoma cell line EMRK1184 treated with WM-1119 and WM-8014 at the doses indicated. **c**, Bioluminescence images of EMRK1184 lymphoma cells expressing luciferase before (day 3) and after (day 14) 11 days of treatment with WM-1119 (50 mg kg⁻¹ four times per day) or PEG400 vehicle control. The red boxes show the regions used for quantification (imaging at days 7, 10 and 12 in Extended Data Fig. 10). **d**, Quantification of the signals measured in all experiments: two cohorts of mice treated with WM-1119 three times per day, combined $n = 6$; two cohorts of mice treated with WM-1119 four times per day, combined $n = 9$; vehicle controls, $n = 15$. One mouse did not respond to WM-1119 treatment, shown in grey. **e**, Dissected spleens obtained after imaging on day 14, taken from the mice shown in c. **f**, Spleen weights of mice treated with WM-1119 or vehicle. n values as stated in d. i.p., intraperitoneal. **g**, Flow-cytometry analysis of spleen cells from vehicle-treated mice and mice treated with WM-1119 (four times per day). The tumour cells were CD19⁺IgM⁻, and normal splenic B cells were CD19⁺IgM⁺. Quantification of flow-cytometry analysis in bone marrow (BM), spleen and peripheral white blood cells (PWBC). $n = 4$ independent experiments for WM-1119 and 2 for WM-8014 in b, and number of mice as indicated in d, f, g in three independent experiments. Data are mean \pm s.e.m. and were analysed by nonlinear regression dose-response curve fit, least squares fit, inhibitor versus response, variable slope (b); one-way ANOVA followed by Bonferroni post hoc test with treatment as the independent factor (d, g), or two-tailed t -tests (f).

(Extended Data Fig. 9), cohorts of mice were injected every 8 h (three times per day, two cohorts of three mice per treatment group) or every 6 h (four times per day, two cohorts of three and six mice per treatment

group; Fig. 4d). Mice were imaged five times over the course of these experiments to monitor the growth of lymphoma. No significant difference between the treatment and control groups was seen before day 10 (Fig. 4d, Extended Data Fig. 10), which was expected as the inhibition of cell proliferation in vitro took approximately seven days. However, by day 14, the cohorts that were treated four times per day with WM-1119 had arrested tumour growth (Fig. 4c, Extended Data Fig. 10), with the exception of one mouse that did not respond (Fig. 4d). Spleen weights in the WM-1119-treatment group (treated four times per day) were substantially lower than spleen weights in the vehicle-treated group, and not significantly different from those of tumour-free eight-week-old mice ($P < 0.0005$ and $P = 0.2$, respectively; Fig. 4e, f). Treatment with WM-1119 three times per day led to a significant reduction in tumour burden and spleen weight, but was not as effective as treatment four times per day (Fig. 4d, f). WM-1119 was well-tolerated; mice showed no generalized ill effects and weight loss was not observed (Extended Data Fig. 10). WM-1119 treatment had no effect on haematocrit, erythrocytes or platelet numbers, but there was overall leukopenia (Extended Data Fig. 10). The proportion and overall number of tumour cells was substantially reduced by WM-1119 treatment (four times per day; Fig. 4g). Analysis by intracellular flow cytometry demonstrated a reduction in H3K9ac in tumour cells ($P = 0.03$; Extended Data Fig. 10). These results demonstrate that WM-1119 is effective in treating lymphoma in vivo.

In summary, using high-throughput screening followed by medicinal chemistry optimization, in-cell assays, biochemical assessment of target engagement and tumour models in mice and fish, we have developed a novel class of inhibitors for a hitherto unexplored category of epigenetic regulators. These inhibitors engage the MYST family of lysine acetyltransferases in primary cells, specifically induce cell cycle exit and senescence, and are effective in preventing the progression of lymphoma in mice.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA-seq data of MEFs treated with WM-8014, WM-2474 and DMSO, of MEFs from *Kat6a*^{-/-} and wild-type embryos and of lymphoma cell line EMRK1184 treated with vehicle and WM-1119 have been submitted to the Gene Expression Omnibus (GEO) database under accession number GSE108244. The crystal structure data for the MYST domain in complex with WM-8014, acetyl-CoA and WM-1119 have been submitted to the Protein Data Bank (PDB) under accession numbers 6BA2, 6BA4 and 6CT2, respectively. Source Data for all graphs are provided.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0387-5>

Received: 10 December 2016; Accepted: 21 June 2018;

Published online 1 August 2018.

1. Lee, K. K. & Workman, J. L. Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* **8**, 284–295 (2007).
2. Allis, C. D. et al. New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633–636 (2007).
3. Voss, A. K. & Thomas, T. MYST family histone acetyltransferases take center stage in stem cells and development. *BioEssays* **31**, 1050–1061 (2009).
4. Katsumoto, T. et al. MOZ is essential for maintenance of hematopoietic stem cells. *Genes Dev.* **20**, 1321–1330 (2006).

5. Thomas, T. et al. Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. *Genes Dev.* **20**, 1175–1186 (2006).
6. Sheikh, B. N. et al. MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells. *Blood* **128**, 2307–2318 (2016).
7. Borrow, J. et al. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat. Genet.* **14**, 33–41 (1996).
8. Huang, F., Abmayr, S. M. & Workman, J. L. Regulation of KAT6 acetyltransferases and their roles in cell cycle progression, stem cell maintenance, and human disease. *Mol. Cell. Biol.* **36**, 1900–1907 (2016).
9. Sheikh, B. N. et al. MOZ (MYST3, KAT6A) inhibits senescence via the INK4A-ARF pathway. *Oncogene* **34**, 5807–5820 (2015).
10. Perez-Campo, F. M. et al. MOZ-mediated repression of *p16^{INK4a}* is critical for the self-renewal of neural and hematopoietic stem cells. *Stem Cells* **32**, 1591–1601 (2014).
11. Sheikh, B. N. et al. MOZ regulates B-cell progenitors and, consequently, *Moz* haploinsufficiency dramatically retards MYC-induced lymphoma development. *Blood* **125**, 1910–1921 (2015).
12. Falk, H. et al. An efficient high-throughput screening method for MYST family acetyltransferases, a new class of epigenetic drug targets. *J. Biomol. Screen.* **16**, 1196–1205 (2011).
13. Holbert, M. A. et al. The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. *J. Biol. Chem.* **282**, 36603–36613 (2007).
14. Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
15. Yan, Z. et al. Cdc6 is regulated by E2F and is essential for DNA replication in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 3603–3608 (1998).
16. Kueh, A. J., Dixon, M. P., Voss, A. K. & Thomas, T. HBO1 is required for H3K14 acetylation and normal transcriptional activity during embryonic development. *Mol. Cell. Biol.* **31**, 845–860 (2011).
17. Voss, A. K., Collin, C., Dixon, M. P. & Thomas, T. Moz and retinoic acid coordinately regulate H3K9 acetylation, *Hox* gene expression, and segment identity. *Dev. Cell* **17**, 674–686 (2009).
18. Voss, A. K. et al. MOZ regulates the *Tbx1* locus, and *Moz* mutation partially phenocopies DiGeorge syndrome. *Dev. Cell* **23**, 652–663 (2012).
19. Chew, T. W. et al. Crosstalk of Ras and Rho: activation of RhoA abates Kras-induced liver tumorigenesis in transgenic zebrafish models. *Oncogene* **33**, 2717–2727 (2014).
20. Adams, J. M. et al. The *c-myc* oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**, 533–538 (1985).

Acknowledgements We thank F. Dabrowski, C. D'Alessandro, WEHI Bioservices, the WEHI FACS laboratory, the MX2 beamline staff at the Australian Synchrotron for their expert help and Z. Gong for the two transgenic zebrafish lines. This work was funded by the Australian Government through NHMRC project grants 1030704, 1080146, Research Fellowships (T.T., A.K.V., G.K.S., J.K.H., M.W.P. and J.B.), the NHMRC IRISS and the Cancer Therapeutics Cooperative Research Centre. The Victorian State Government OIS Grants to WEHI, Monash and St Vincent's Institute are gratefully acknowledged.

Reviewer information Nature thanks P. Adams, R. Marmorstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.T. was responsible for initiating the project. T.T. and A.K.V. supervised the project, performed experiments and wrote the manuscript. Medicinal chemistry: supervised by J.B.B., team: D.J.L., N.N., B.C. and H.R.L. Structural biology: S.J.H., M.C.C., B.R., T.S.P. and M.W.P. Chemical screening, protein biochemistry and assays: M.d.S., J.B., P.P., M.H., O.D., M.L.D., H.F., I.P.S. and B.J.M. Pharmacology: S.A.C. and K.L.W. Bioinformatics: G.P., A.L.G. and G.K.S. Cell-based assays, molecular biology and biochemistry: N.L.D., J.W., H.M.M., Y.Y., H.K.V., M.I.B., R.E.M., B.K.D., B.W., N.Z., S.W., B.N.S. and B.J.A. Zebrafish model: S.M., K.J.M., A.J.S., K.D. and J.K.H. Mouse cancer models: G.L.K., M.S.B., J.R., A.N., E.D.H., R.W.J., N.D.H. and A.S.

Competing interests The authors declare no competing interests.

Additional information

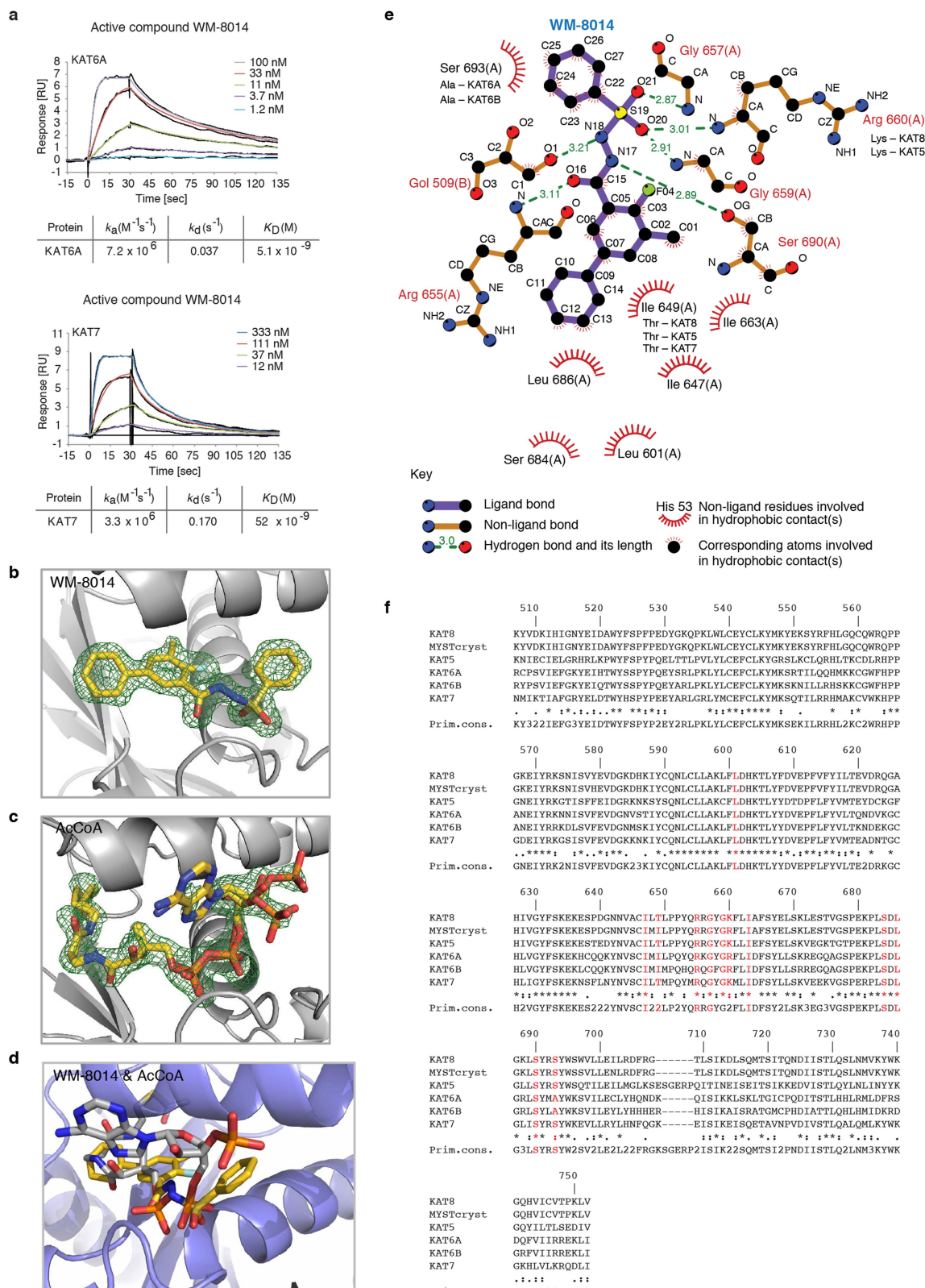
Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0387-5>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0387-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.B.B., A.K.V. or T.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

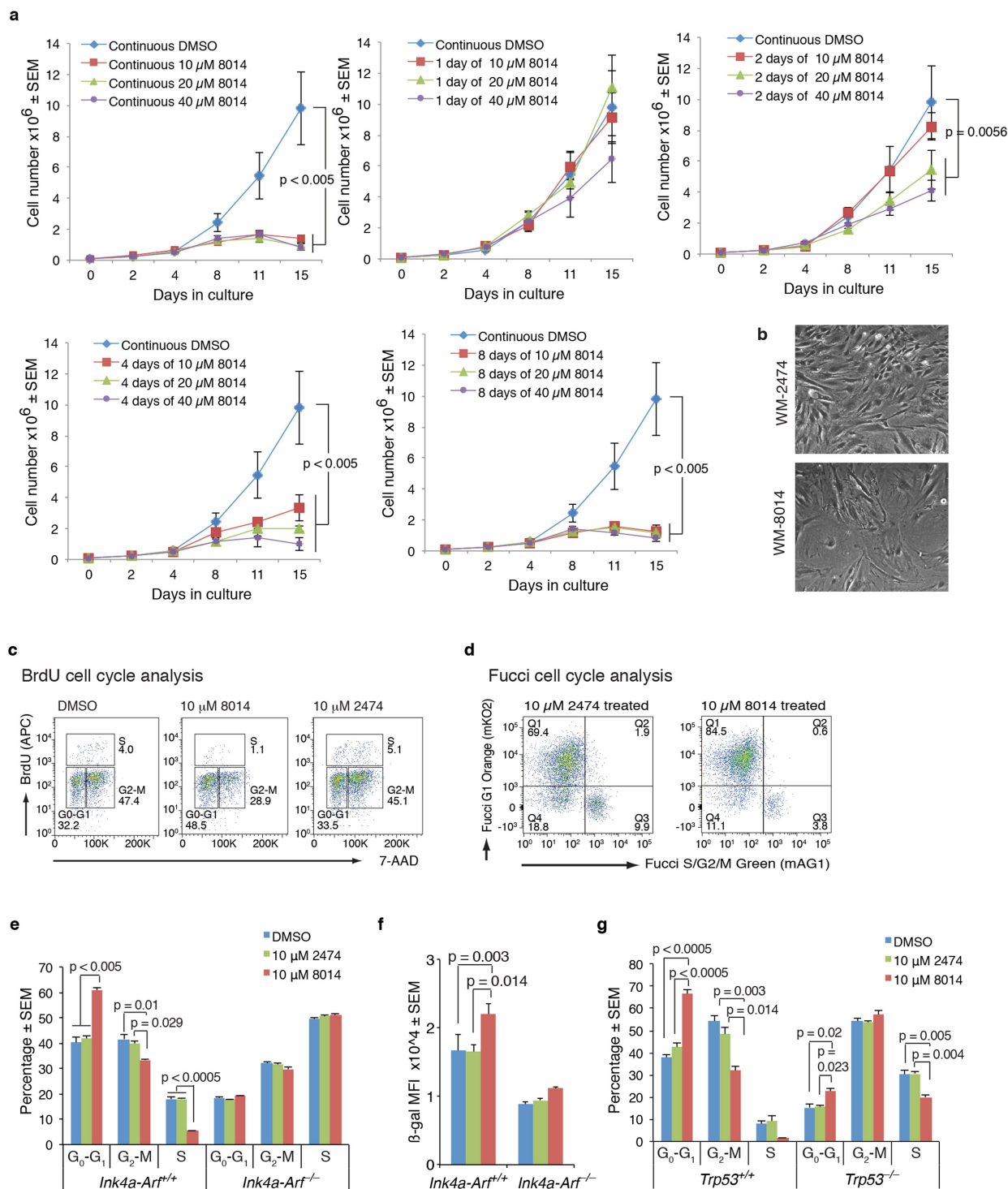


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Binding characteristics of the MYST domain–WM-8014 protein–ligand interaction and comparison of MYST family histone acetyltransferase domains. **a**, SPR binding data for the interaction of WM-8014 with immobilized KAT6A and KAT7 MYST domains. Injected concentrations of WM-8014 are indicated. Binding responses (data; black sensorgrams) are overlaid, fitted curves of a 1:1 kinetic interaction model that included mass transport component (coloured lines), as well as derived kinetic rate constants (k_a , k_d) and equilibrium dissociation (K_D) constant. One of at least two experiments is shown. **b**, WM-8014 bound to MYST^{Cryst}, with the WM-8014 OMIT electron density map contoured to 3σ shown in green. **c**, Acetyl-CoA bound to MYST^{Cryst}, with the acetyl-CoA OMIT map contoured to 3σ shown in green. **d**, Ribbon diagram showing WM-8014 and acetyl-CoA superimposed. **e**, Protein–ligand interactions (LIGPLOT)²¹ between

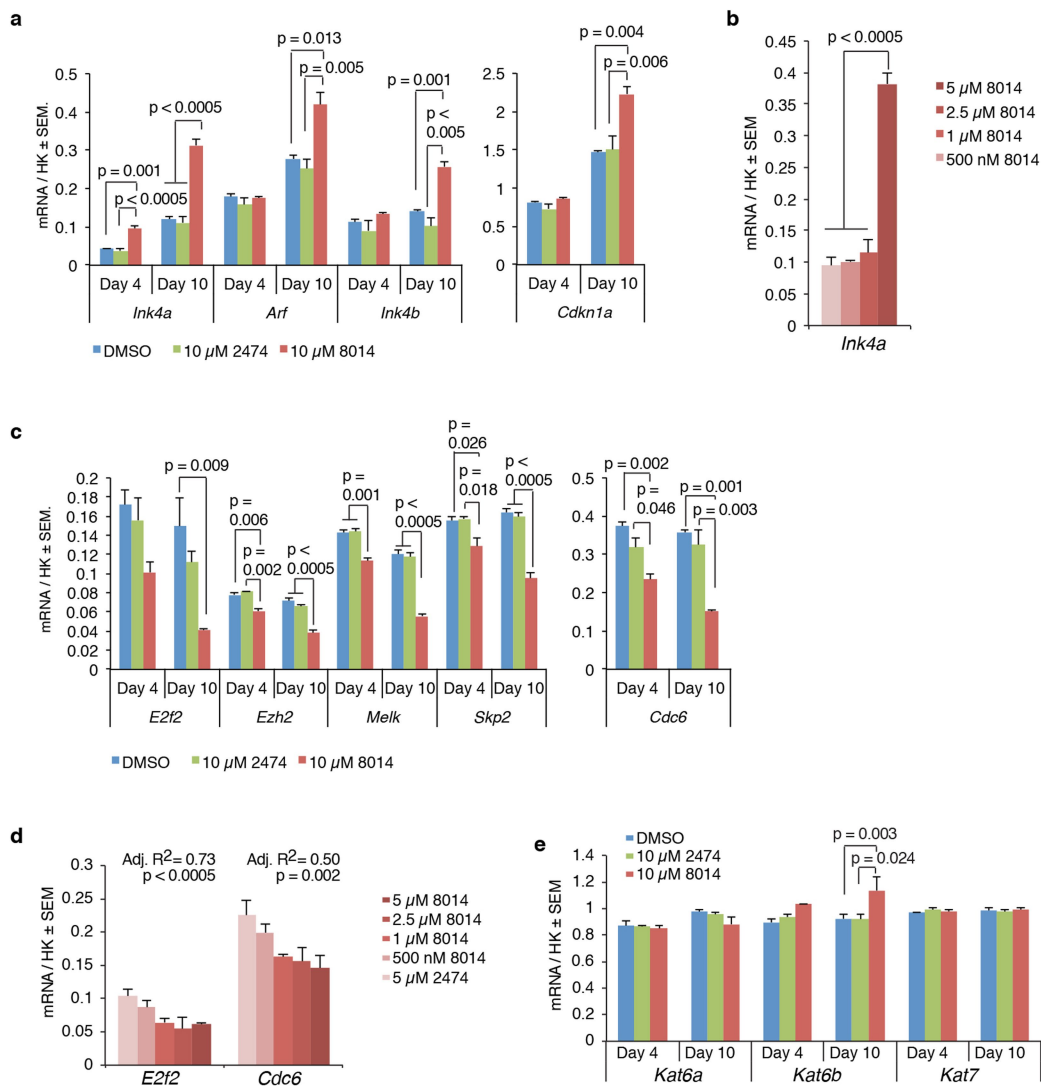
WM-8014 and amino acids within the acetyl-CoA-binding pocket of the MYST domain. The amino acids that differ between MYST family members are indicated. Data collection and refinement statistics of the WM-8014 and acetyl-CoA co-crystal structures can be found in Supplementary Table 4. The overall structure of WM-8014 bound to MYST^{Cryst} is nearly identical to the MYST^{Cryst}–acetyl-CoA complex. The pantothenate arm of acetyl-CoA adopts an identical position to published MYST HAT domain structures; as observed previously, there are differing positions for the 3'-phosphate ADP¹³. Autoacetylation of K604 was observed, as expected²². Gol denotes glycerol. **f**, Comparison of the conserved MYST domain between MYST family proteins. MYST^{Cryst} is a MYST domain modified to improve solubility and used in crystallization studies. Numbering as in KAT6A sequence, NP_006757.2; amino acids interacting with WM-8014 (depicted in the LIGPLOT) are shown in red.

21. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
22. Yuan, H. et al. MYST protein acetyltransferase activity requires active site lysine autoacetylation. *EMBO J.* **31**, 58–70 (2012).



Extended Data Fig. 2 | Time course of MEF growth inhibition upon treatment with WM-8014, and requirement for INK4A/ARF and p53 for WM-8014-induced cell cycle arrest. a, MEF proliferation after treatment with three high concentrations of WM-8014. MEFs were treated either continuously for 15 days, or treatment was discontinued after 1, 2, 4 or 8 days to determine whether cells could re-enter the cell cycle. **b**, Phase-contrast images of MEFs after 15 days of treatment with 10 μM WM-8014 or 10 μM WM-2474. Note cells with senescence morphology; that is, large nuclei indicating endoreplication without cell division and extensive cytoplasm (WM-8014 panel). **c**, Flow-cytometry gating strategy for the cell cycle analysis using incorporation of the nucleotide analogue BrdU to mark cells in S phase and 7-aminoactinomycin D (7-AAD) to determine 2N (G0/G1) and 4N (G2/M) DNA content. **d**, Flow-cytometry gating strategy for the cell cycle analysis of transgenic Fucci cells that express Azami Green in mid-S phase, G2 and M, Kusabira Orange in mid-late G1,

are double-positive yellow in early S phase and double-negative in early G1. **e**, Cell cycle analysis of *Cdkn2a* null (*Ink4a^{-/-}Arf^{-/-}*) and littermate control cells after treatment for 8 days with WM-8014, vehicle and the inactive compound WM-2474. MEFs were exposed to BrdU for 1 h before flow-cytometry analysis of BrdU incorporation during DNA synthesis (S phase) and DNA content of 2N (G0/G1) compared with 4N (G2/M) using 7-AAD. **f**, Senescence-associated β -galactosidase activity in *Cdkn2a^{-/-}* and control MEFs after treatment for 15 days with 10 μM WM-8014, 10 μM WM-2474 or DMSO vehicle control. **g**, Cell cycle analysis of *Trp53* null MEFs (*Trp53^{-/-}*) and littermate control cells after treatment with WM-8014, vehicle and inactive compound WM-2474, as in **c**. $n = 3$ MEF isolates per genotype (**a–e**). Data are mean \pm s.e.m., and were analysed by two-way ANOVA within duration of treatment with concentration and days of culture as the independent factors (**a**), or by one-way ANOVA followed by Bonferroni post hoc test (**e–g**).

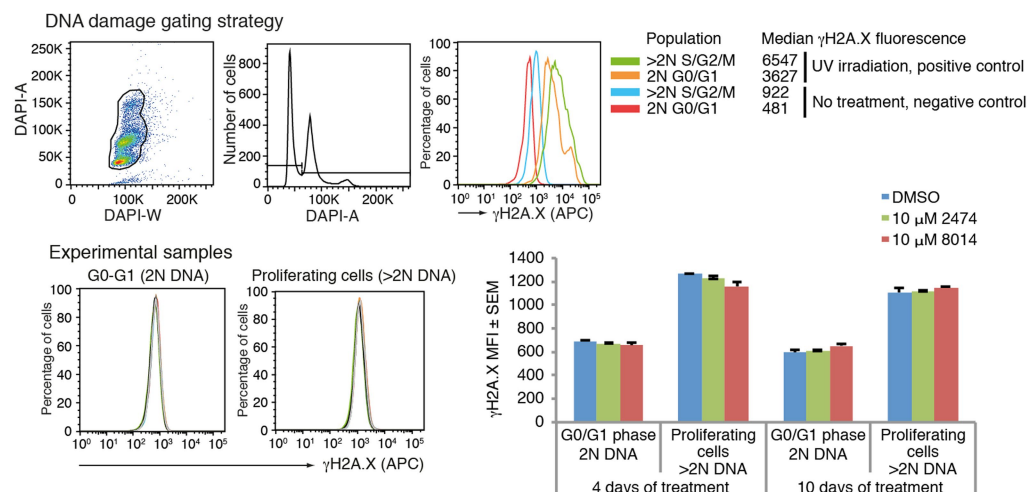
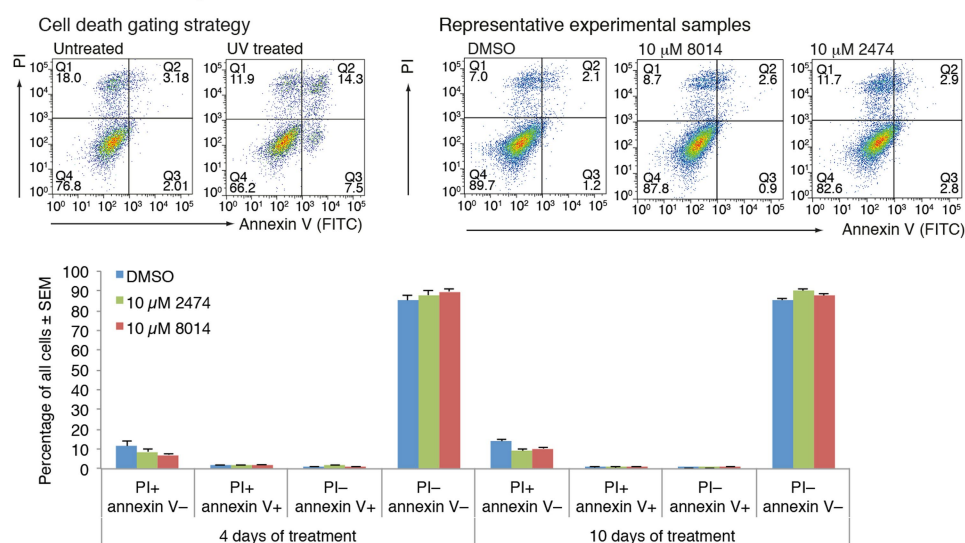


Extended Data Fig. 3 | The effect of WM-8014 on cell proliferation is mediated through the cell cycle regulators p16^{INK4A} and p19^{ARF}.

a, RT-qPCR analysis of expression levels of cell cycle regulators *Ink4a* and *Arf* (alternative splice products of the *Cdkn2a* locus), *Ink4b* (also known as *Cdkn2b*) and *Cdkn1a* (encoding p21^{WAF1/CIP1}) mRNA in MEFs treated for 4 days and 10 days with 10 μ M WM-8014 or 10 μ M control WM-2474.

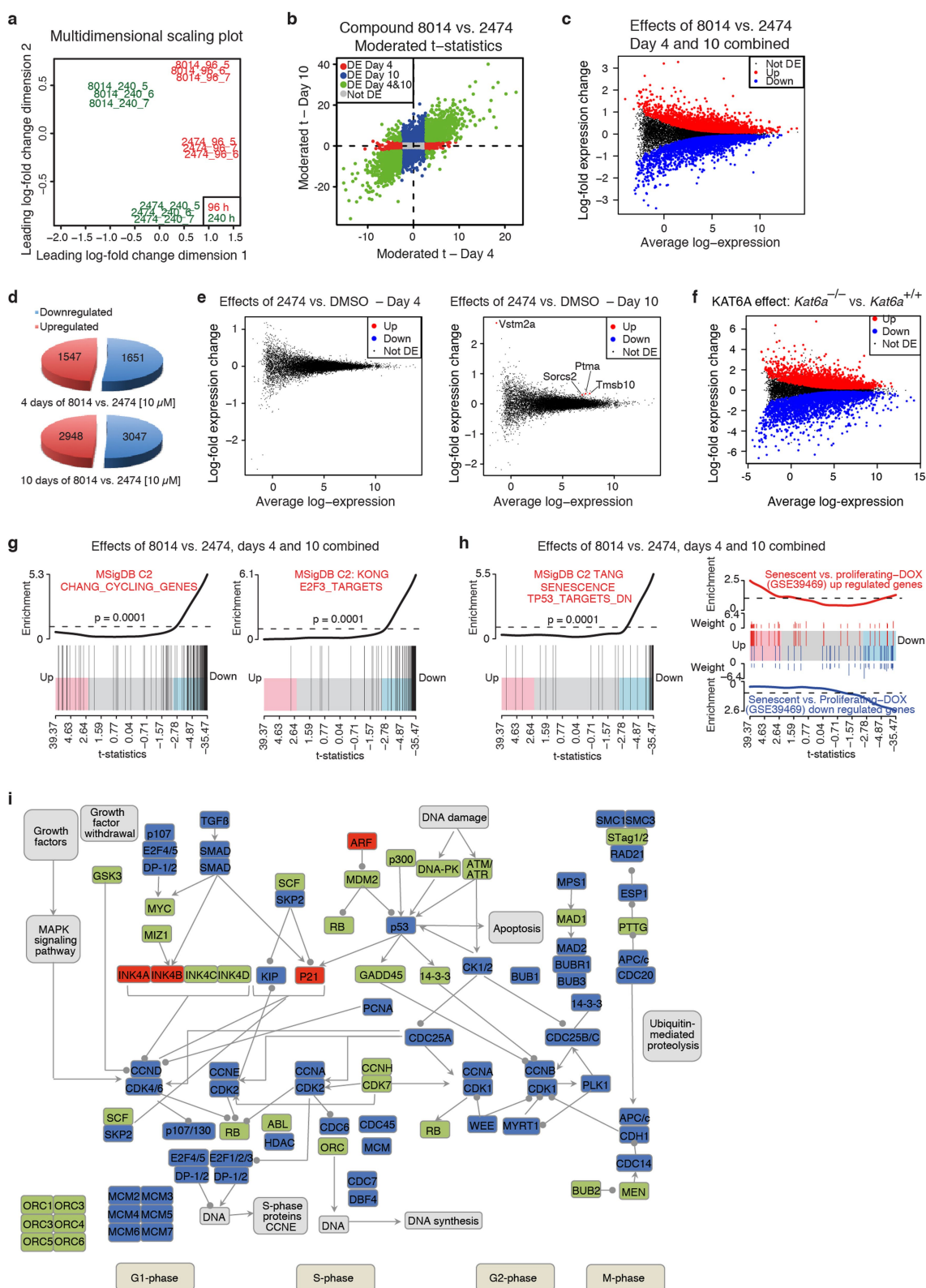
b, Dose-response plots of WM-8014 induction of *Ink4a* mRNA expression in MEFs. **c**, RT-qPCR analysis of expression changes in the KAT6A target gene detected by RNA-seq. MEFs were treated for 4 days and 10 days with

10 μ M WM-8014, 10 μ M control WM-2474 or DMSO. **d**, Dose-response plots of WM-8014-dependent reduction in *E2f2* and *Cdc6* mRNA levels in MEFs. **e**, Levels of mRNA coding for MYST-family proteins after treatment of MEFs for 4 days or 10 days with WM-8014, vehicle or the inactive compound WM-2474. $n = 3$ MEF isolates treated with WM-8014, WM-2474 or vehicle (**a-e**). Data are mean \pm s.e.m. and are analysed by one-way ANOVA followed by Bonferroni post hoc test (**a-c, e**) and by regression analysis (**d**). mRNA levels normalized to housekeeping genes (HK) were regressed on the log(concentration) of WM-8014 (**d**).

a DNA damage analysis**b** Cell death analysis

Extended Data Fig. 4 | WM-8014 causes cell cycle exit and senescence in MEFs, but not DNA damage or cell death. **a**, Assessment of DNA damage using flow cytometry to detect γ H2A.X. Top, exposure of MEFs to ultraviolet light (positive control). Bottom, experimental samples. Quantification is displayed in the bar graph. **b**, Flow-cytometry gating strategy for cell death analysis and representative experimental samples. Negative and positive controls (untreated and ultraviolet-light-irradiated

cells, respectively) are shown in the left panels. Annexin V marks phosphatidylserine externalization on cells undergoing apoptosis, propidium iodide (PI) uptake marks cells undergoing other forms of cell death, annexin V/PI double-positive staining marks cells in late-stage apoptosis. $n = 3$ cultures (**a**, **b**). Data are mean \pm s.e.m. and were analysed by one-way ANOVA with treatment as the independent factor.

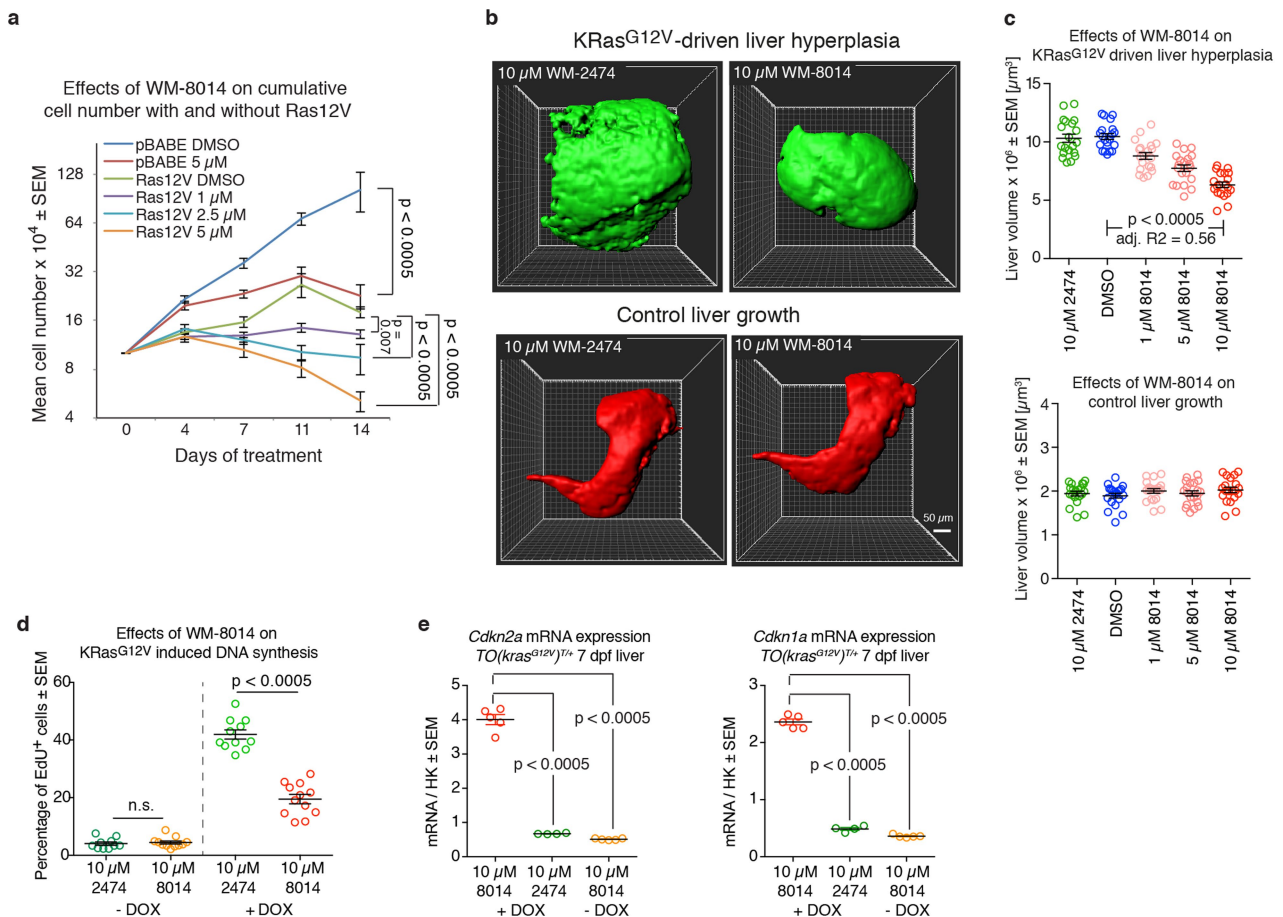


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | WM-8014 treatment induces a gene signature of cellular senescence. **a**, Multidimensional scaling plot (\log_2 fold changes) showing clustering of MEF expression profiles after treatment with WM-8014 or control WM-2474. MEFs were isolated from 3 different embryos, numbered 5, 6 and 7 and treated for 4 days (96 h, red) or 10 days (240 h, green). **b**, Scatter plot showing gene-wise t -statistics for differentially expressed (DE) genes ($\text{FDR} < 0.05$) between the compounds at day 4 and day 10. Most genes were equally affected by 4 days or 10 days of treatment (green). Genes differentially expressed at day 10 only are highlighted blue, those differentially expressed at day 4 only are highlighted red. **c**, Mean-difference plot of treatment, \log_2 fold changes versus average \log_2 expression. Treatment effects at 4 days and 10 days have been averaged. Differentially expressed genes are highlighted in red or blue as indicated ($\text{FDR} < 0.05$). **d**, Number of differentially expressed genes for MEFs treated with WM-8014 versus WM-2474 ($\text{FDR} < 0.05$). **e**, Mean-difference plot of treatment \log_2 fold changes versus average \log_2 expression comparing WM-2474 to vehicle DMSO. The four differentially expressed genes ($\text{FDR} < 0.05$) are marked in red. **f**, Mean-difference plot of \log_2 fold changes versus average \log_2 expression comparing *Kat6a*^{-/-}

MEFs with *Kat6a*^{+/+} control MEFs. Differentially expressed genes are highlighted in red or blue as indicated ($\text{FDR} < 0.05$). **g**, Genes typical of cycling cells²³ and E2F3 target genes²⁴ are downregulated in MEFs treated with WM-8014 versus WM-2474 (combined day 4 and day 10 treatment; ROAST gene set tests, $P = 0.0001$). **h**, Genes downregulated during p53-induced cellular senescence²⁵ were downregulated in MEFs treated with WM-8014 versus WM-2474 (combined day 4 and day 10 treatment; ROAST $P = 0.0001$). Differentially expressed genes in cellular senescence²⁶ are strongly correlated with differentially expressed genes in MEFs treated with WM-8014 versus WM-2474 (ROAST $P = 0.0039$). **i**, DAVID²⁷ was used to test for functional enrichment in genes downregulated after treatment with WM-8014 versus WM-2474, with $\text{FDR} < 0.05$. Cell cycle regulation was the top enriched pathway ($\text{FDR} = 1.58 \times 10^{-16}$), with 85% of the genes downregulated after 10 days of treatment with WM-8014. Schematic drawing is based on mmu04110: cell cycle²⁸. Downregulated genes are shaded blue; unchanged, green; upregulated genes (*Ink4a*, *Arf*, *Ink4b* and *p21*) are shaded red. Data were collected from $n = 3$ MEFs isolates from 3 different embryos per treatment group, WM-8014 or WM-2474 treatment, for 96 h or 240 h.

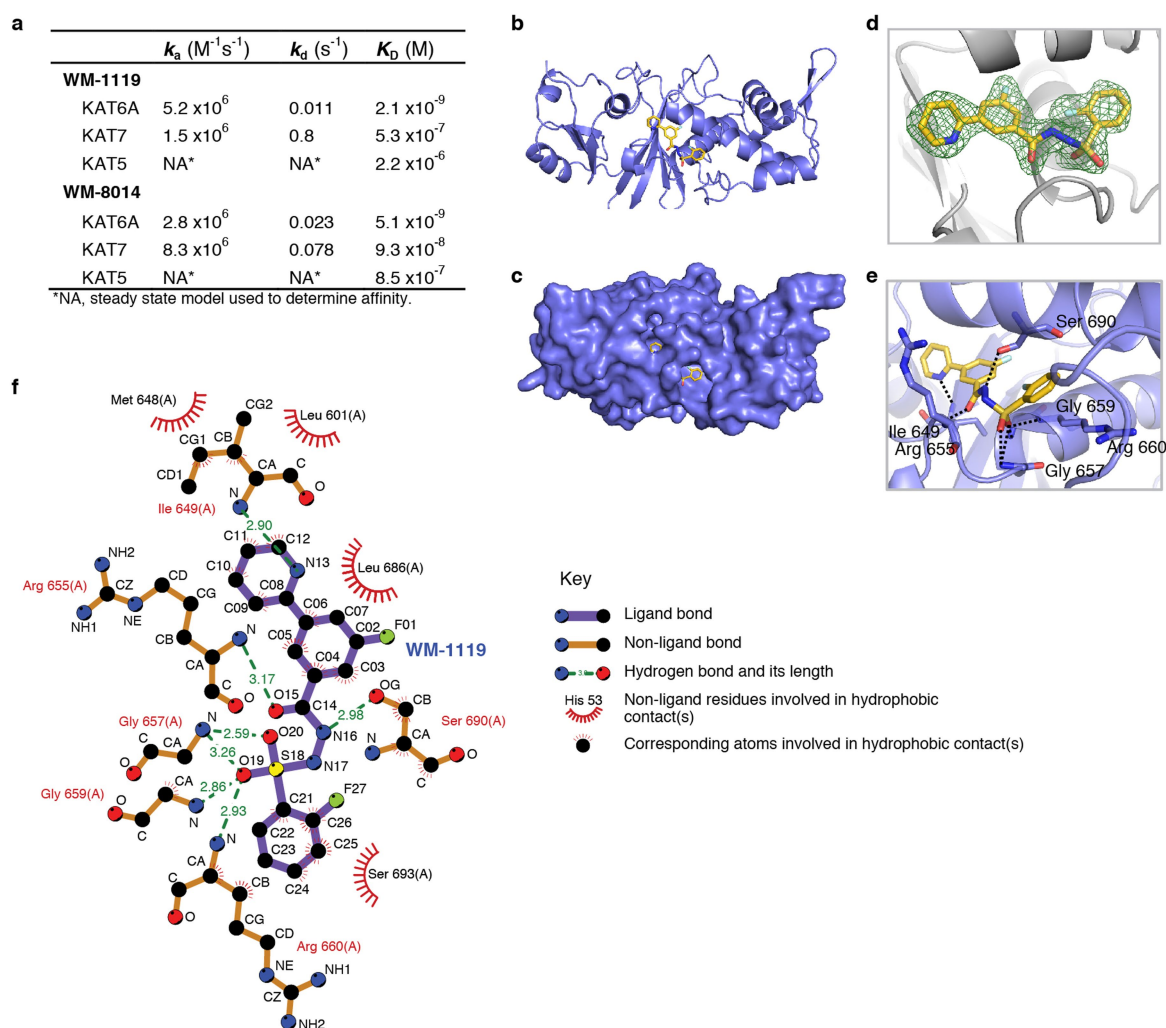
23. Chang, H. Y. et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, e7 (2004).
24. Kong, L. J., Chang, J. T., Bild, A. H. & Nevins, J. R. Compensation and specificity of function within the E2F family. *Oncogene* **26**, 321–327 (2007).
25. Tang, X., Milyavsky, M., Goldfinger, N. & Rotter, V. Amyloid- β precursor-like protein APLP1 is a novel p53 transcriptional target gene that augments neuroblastoma cell death upon genotoxic stress. *Oncogene* **26**, 7302–7312 (2007).
26. Lujambio, A. et al. Non-cell-autonomous tumor suppression by p53. *Cell* **153**, 449–460 (2013).
27. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
28. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).



Extended Data Fig. 6 | WM-8014 potentiates oncogene-induced senescence. **a**, Growth curves of MEFs expressing empty vector control (pBABA) or oncogenic²⁹ HRAS^{G12V} treated with increasing concentrations of WM-8014 as indicated or DMSO vehicle control. All experiments were performed in 3% O₂. **b**, The effects of WM-8014 treatment in a zebrafish model of hepatocellular carcinoma¹⁹. Doxycycline-inducible, liver-specific expression of a GFP-kras^{G12V} transgene leads to the accumulation of a constitutively active, GFP-tagged form of KRAS in hepatocytes. TO-kras^{G12V} transgenic embryos were treated with doxycycline at 2 days post fertilization (dpf) and 5 dpf to initiate KRAS^{G12V}-driven hepatocyte proliferation. The size of the liver was measured by two-photon microscopy. Representative three-dimensional reconstructions of whole livers from image stacks after treatment of transgenic zebrafish Tg(TO-kras^{G12V}) expressing KRAS^{G12V} and GFP (green) in the liver or

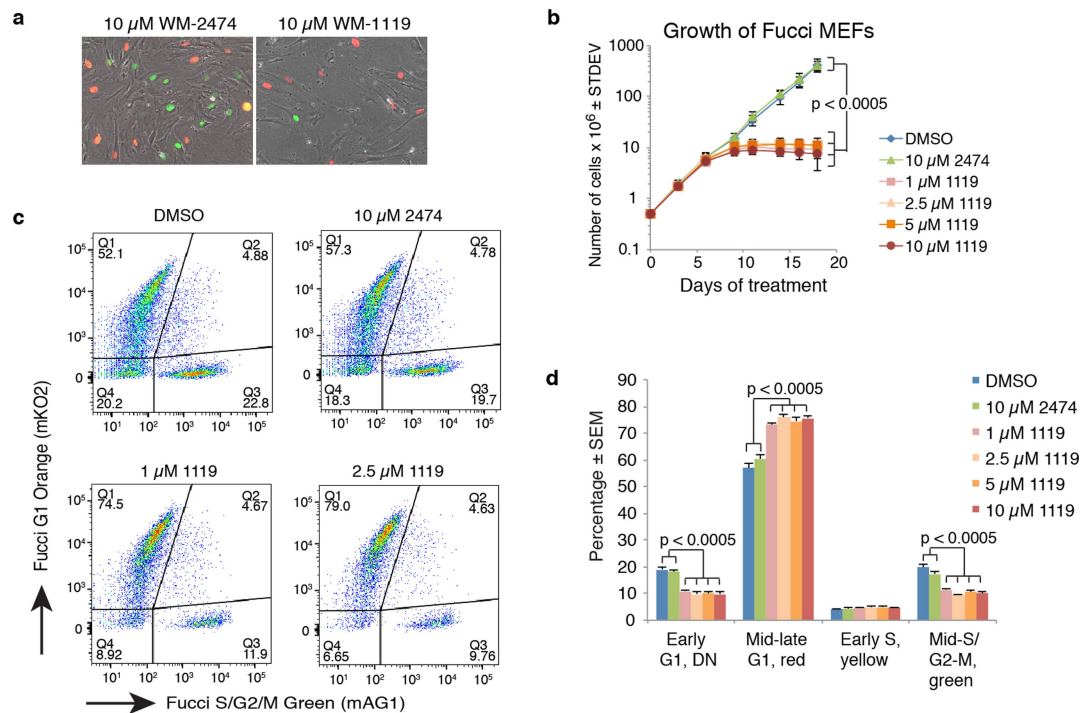
transgenic zebrafish Tg(lfabp10:RFP;elaA:eGFP) expressing only RFP (red). **c**, Quantification of liver volume. **d**, Incorporation of the nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU) after treatment of transgenic zebrafish expressing KRAS^{G12V} or control zebrafish with WM-8014 or control compound WM-2474. **e**, RT-qPCR determination of *Cdkn2a* (*Ink4a*) and *Cdkn1a* (encoding p21^{WAF1/CIP1}) mRNA levels in transgenic zebrafish Tg(TO-kras^{G12V}) treated as described in **b**. $n = 6$ independent cultures (**a**), 20 zebrafish (**b, c**), 10–12 zebrafish (**d**) and 4–5 zebrafish (**e**). Data are mean \pm s.e.m. and were analysed by two-way ANOVA (**a**) or one-way ANOVA (**d, e**) followed by Bonferroni post hoc test with treatment and with or without treatment duration as the independent factors or by linear regression analysis regressing liver volume on WM-8014 concentration (**c**).

29. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).



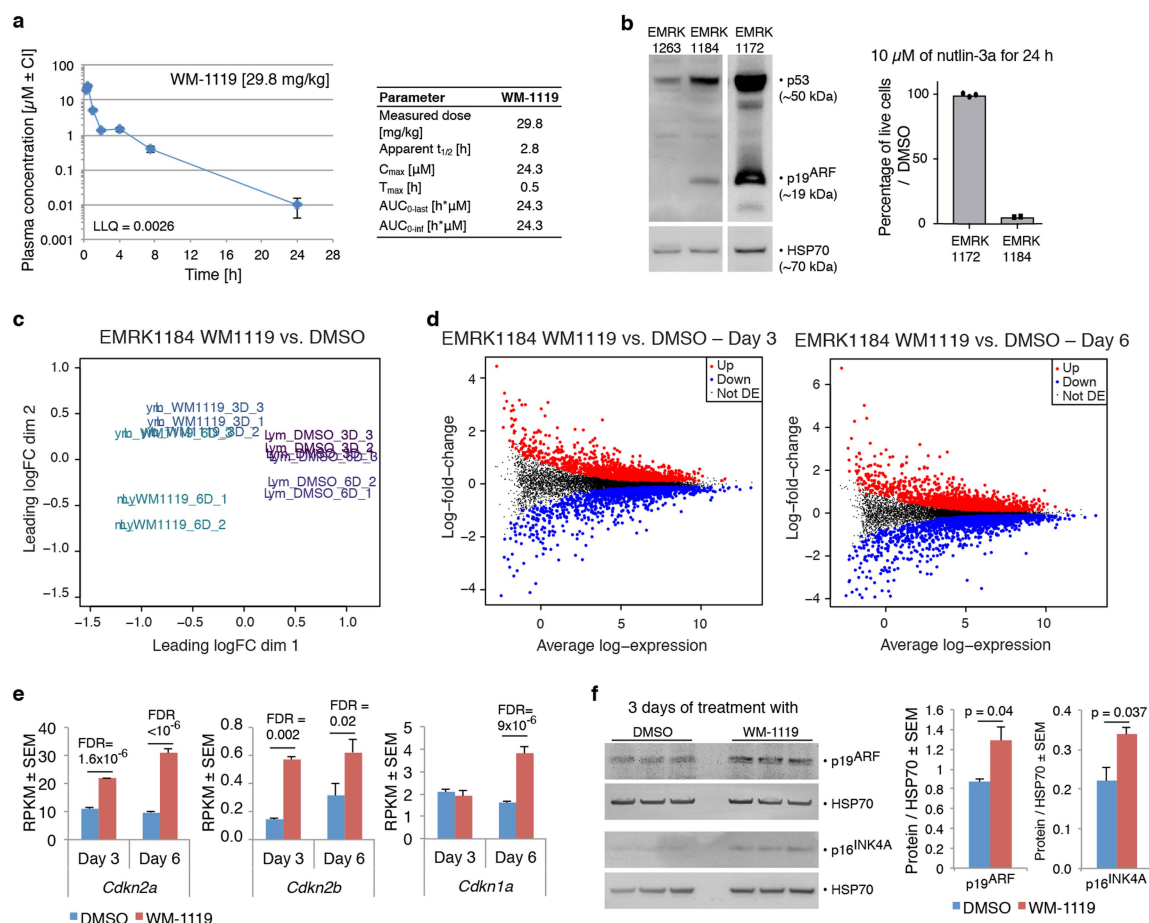
Extended Data Fig. 7 | Medicinal chemistry optimization of WM-8014, designed to reduce plasma-protein binding, resulted in compound WM-1119. a, SPR binding data for the interaction of WM-1119 with immobilized KAT6A, KAT7 and KAT5, compared with the interaction of WM-8014. **b**, Crystal structure of WM-1119 bound to the MYST lysine acetyltransferase domain (MYST^{Cryst}). Ribbon diagram of MYST^{Cryst} (blue) showing WM-1119 (yellow, with element colouring) bound to the acetyl-CoA-binding site. Data collection and refinement statistics of the WM-1119 co-crystal structures (2.13 Å resolution) are listed in

Supplementary Table 5. PDB: 6CT2. **c**, Space-filling model showing WM-1119 in the acetyl-CoA-binding pocket of MYST^{Cryst}. **d**, WM-1119 bound to MYST^{Cryst} with the OMIT electron density map contoured to 3σ shown in green. **e**, Ribbon diagram of MYST^{Cryst} showing key amino acids interacting with WM-1119, in stick fashion with element colouring. Hydrogen bonds are shown as dashed lines. **f**, Schematic diagram of protein-ligand interactions (LIGPLOT)²¹ showing interactions between the compound WM-1119 and amino acids within the acetyl-CoA-binding pocket of the MYST domain derived from the crystal structure.



Extended Data Fig. 8 | WM-1119 causes retention of cells in the G1 phase of the cell cycle. **a**, WM-1119 causes cell cycle arrest in MEFs grown in 3% O₂. Epifluorescence phase contrast images of Fucci MEFs after 8 days of treatment with 10 μ M WM-1119 compared to 10 μ M control WM-2474-treated cells. **b**, WM-1119 was tested at concentrations from 1 to 10 μ M, compared to DMSO or 10 μ M inactive compound WM-2474. The cell number under each condition was assessed at passage. **c**, Flow-cytometry analysis of Azami Green (mAG1; mid-S, G2, M), Kusabira Orange (mKO2; mid-late G1), double-positive yellow (early S)

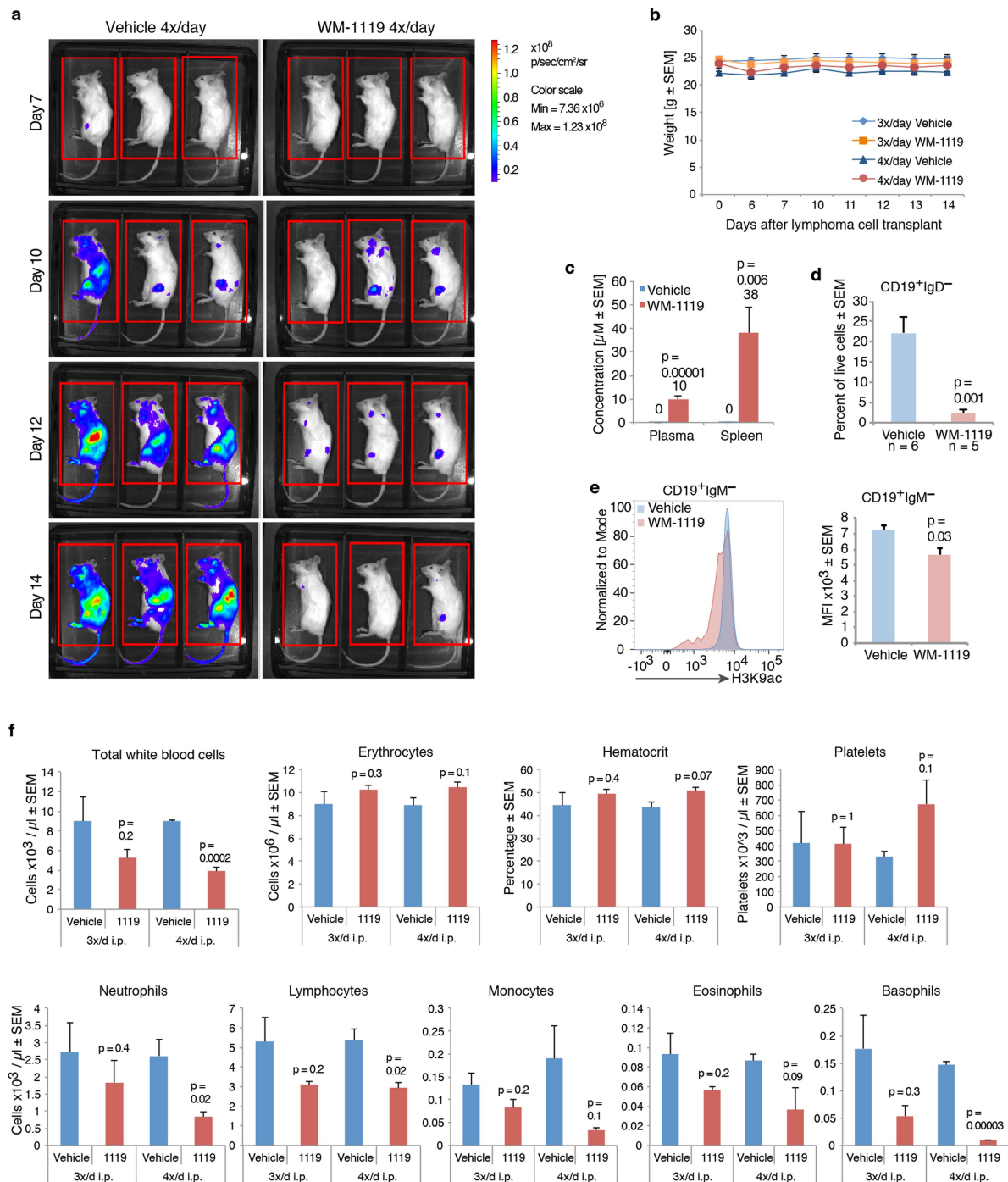
and double-negative (DN, early G1). Dot plots are shown for DMSO and 10 μ M WM-2474 control treatment groups, and after treatment with 1 μ M and 2.5 μ M active compound WM1119. **d**, Percentage of cells in each phase of the cell cycle, quantified for all treatment groups. A higher proportion of WM-1119-treated cells is in mid-late G1. $n = 3$ independent MEF isolates. Data are mean \pm s.e.m. and were analysed by two-way ANOVA (**b**) or one-way ANOVA followed by Bonferroni post hoc test (**d**) with treatment and with or without time as the independent factors.



Extended Data Fig. 9 | Characterization of WM-1119 and lymphoma cell line EMRK1184.

a, Pharmacokinetic parameters for WM-1119 in mice following intraperitoneal injection. Note that the plasma concentration falls below 1 µM after 4 h. Data of $n = 2$ mice are shown. **b**, Characterization of the *Eµ-Myc* lymphoma cell line EMRK1184. Left, Western blot of p53 and p19^{ARF}. The negative control cell line EMRK1263 lacks the ARF (p19^{ARF}) band. Upregulation of p53 protein levels in positive control cell line EMRK1172 indicates non-functional p53 (commonly mutations in the DNA-binding domain). Right, EMRK1184 cells were sensitive to nutlin-3a-induced cell death, indicating intact p53. By contrast, EMRK1172 cells are insensitive to nutlin-3a. p53 exon sequencing of EMRK1184 using the MiSeq system (Illumina) confirmed wild-type p53 exon sequences. **c**, Multidimensional scaling plot showing two-dimensional clustering of the EMRK1184 lymphoma cell culture expression profiles. EMRK1184 lymphoma cells were treated for either 3 days or 6 days, in triplicate, with WM-1119 or vehicle before RNA-seq. Distances on the plot corresponding to leading log₂ fold change between

gene expression profiles. **d**, Mean-difference plot of treatment log₂ fold changes versus average log₂ expression for gene expression changes in the EMRK1184 lymphoma cell line after treatment for 3 days and 6 days with WM-1119 or DMSO vehicle control. Differentially expressed genes are highlighted (FDR < 0.05). **e**, mRNA levels assessed by RNA-seq of EMRK1184 cells treated with WM-1119 or vehicle. mRNA levels for *Cdkn2a* (coding for p16^{INK4A}/p19^{ARF}), *Cdkn2b* and *Cdkn1a* are shown. **f**, Western blot and densitometry analysis showing p16^{INK4A} and p19^{ARF} protein in EMRK1184 treated with WM-1119 or vehicle for 3 days. Each lane represents one independent culture, a total of 6 lanes (= 6 cultures) are shown. Data are mean ± confidence interval (**a**) or ± s.e.m. (**e**, **f**). Data in **b** were derived from three (EMRK1172) and two (EMRK1184) independent cell culture experiments, reflected by the individual data points. Data in **c–e** were derived from three independent cultures per treatment group and analysed as described under RNA-seq in the Supplementary Methods. Data in **f** were analysed by one-way ANOVA followed by Bonferroni post hoc test.



Extended Data Fig. 10 | WM-1119 is effective in inhibiting tumour progression. a, Tumour development monitored by luciferase activity and bioluminescence imaging. Lateral images of mice treated four times per day with either vehicle or WM-1119 between day 7 and day 14 after injection with tumour cells. Baseline tumour burden is shown at higher sensitivity setting for day 3 (before treatment) in Fig. 4. Here, images at days 7, 10, 12 and 14 after tumour cell transplant are shown on the same, less-sensitive scale. Mice are imaged in the same order. Red boxes indicate the area used for quantification. **b**, Mouse body weights are not affected by treatment three or four times per day. **c**, Concentration of WM-1119 in peripheral blood and spleen 6 h after the final injection (four times per day; $n = 6$ mice per treatment group). **d**, Flow-cytometry analysis of total spleen cells from vehicle- or WM-1119-treated groups (four times per day; analysis of spleens assayed in **a** to identify tumour cells independently of luciferase expression). The lymphoma cell line EMRK1184 has a cell surface phenotype of CD19⁺IgM⁻IgD⁻. Flow cytometry was used to quantify the CD19⁺IgD⁻ population; this can be distinguished from

normal splenic B cell populations, which are CD19⁺IgD⁺. **e**, Intracellular flow-cytometry analysis of H3K9ac in tumour cells. Left, the histogram shows H3K9ac levels in the remaining tumour cells (CD19⁺IgM⁻) in spleens of the WM-1119-treated mice (red profile) compared to the vehicle-treated mice (blue profile). The shift in the red (WM-1119-treated) profile compared to the blue (vehicle-treated) profile indicates a reduction in signal. Right, the median fluorescence intensity (mean \pm s.e.m.) is shown in the bar graph. **f**, Peripheral blood analysis of vehicle- or WM-1119-treated mice. The cohort of mice that was treated three times per day is compared to the cohort that was treated four times per day. Images representative of $n = 9$ mice per treatment group in the four-times-per-day treatment regime (**a**). $n = 3$ mice per treatment group (**b**, **d**–**f**) and $n = 6$ mice per treatment group in (**c**). Data are mean \pm s.e.m. and were analysed by one-way ANOVA with treatment as the independent factor followed by Bonferroni post hoc test (**b**), or two sided *t*-test (**c**, **d**, **f**) or one-sided *t*-test (**e**).

Cryo-EM of the dynamin polymer assembled on lipid membrane

Leopold Kong¹, Kem A. Sochacki², Huaibin Wang¹, Shunming Fang^{1,4}, Bertram Canagarajah¹, Andrew D. Kehr¹, William J. Rice³, Marie-Paule Strub², Justin W. Taraska² & Jenny E. Hinshaw^{1*}

Membrane fission is a fundamental process in the regulation and remodelling of cell membranes. Dynamin, a large GTPase, mediates membrane fission by assembling around, constricting and cleaving the necks of budding vesicles¹. Here we report a 3.75 Å resolution cryo-electron microscopy structure of the membrane-associated helical polymer of human dynamin-1 in the GMPPCP-bound state. The structure defines the helical symmetry of the dynamin polymer and the positions of its oligomeric interfaces, which were validated by cell-based endocytosis assays. Compared to the lipid-free tetramer form², membrane-associated dynamin binds to the lipid bilayer with its pleckstrin homology domain (PHD) and self-assembles across the helical rungs via its guanine nucleotide-binding (GTPase) domain³. Notably, interaction with the membrane and helical assembly are accommodated by a severely bent bundle signalling element (BSE), which connects the GTPase domain to the rest of the protein. The BSE conformation is asymmetric across the inter-rung GTPase interface, and is unique compared to all known nucleotide-bound states of dynamin. The structure suggests that the BSE bends as a result of forces generated from the GTPase dimer interaction that are transferred across the stalk to the PHD and lipid membrane. Mutations that disrupted the BSE kink impaired endocytosis. We also report a 10.1 Å resolution cryo-electron microscopy map of a super-constricted dynamin polymer showing localized conformational changes at the BSE and GTPase domains, induced by GTP hydrolysis, that drive membrane constriction. Together, our results provide a structural basis for the mechanism of action of dynamin on the lipid membrane.

Dynamin family members are mechanochemical GTPases that catalyse membrane remodelling during essential cellular processes¹. Mutations in dynamins are associated with neuropathies⁴ and atypical expression of dynamins is associated with diverse cancers⁵, while several viruses (for example, HIV) hijack dynamin-dependent pathways^{6,7}. All dynamins are elongated, modular proteins, sharing a structurally conserved N-terminal GTPase domain connected to a four-helix stalk by a three-helix BSE⁸. The prototypical member, dynamin, also contains the lipid-binding PHD and a proline/arginine-rich domain (PRD) that interacts with dynamin partners that contain an SRC homology 3 (SH3) domain⁹. Crystal structures have suggested that, in the absence of lipid, dynamin exists as a homo-tetramer formed from two dimers². The dimer is held together by an extensive interface at the stalk domain (interface 2)^{10,11}, whereas the tetramer is stabilized at the junction between the stalk and the BSE (interface 1) and at the membrane-facing end of the stalks (interface 3)² (Fig. 1). In all crystal structures, the PHD is either disordered or tucked up into its own stalk. In the assembled state, at the necks of budding vesicles or bound to lipid *in vitro*, low-resolution cryo-electron microscopy (cryo-EM) structures have suggested that dynamin further oligomerizes into a helical polymer encasing a lipid tube with an additional GTPase domain dimer interface (interface G2) between the rungs of the helix³. When it binds and hydrolyses GTP, the helical polymer constricts the underlying

membrane from a thick (more than 20 nm) inner lumen diameter down to a hemi-fission state with a diameter below 3.4 nm^{12,13} and catalyses membrane fission. Although these points are well established, the function of GTP energy in relation to membrane constriction and fission and the molecular details of the membrane-bound conformations remain unknown for biologically relevant forms of dynamin¹⁴.

To provide a structural basis for the mechanism by which dynamin acts, we determined a 3.75 Å cryo-EM map of the constricted dynamin-1 (dyn) polymer lacking the intrinsically disordered PRD, assembled on lipid and treated with the non-hydrolysable GTP analogue GMPPCP (dyn^{GMPPCP}) (Fig. 1a). We have complemented this structure with a cryo-EM reconstruction of a 10.1 Å resolution super-constricted dynamin polymer treated with GTP (dyn^{GTP}) (Extended Data Figs. 1, 2). Whereas dyn^{GMPPCP} represents the GTP-bound form of the dynamin polymer, dyn^{GTP} may constitute an intermediate conformation between GTP binding and GTP hydrolysis.

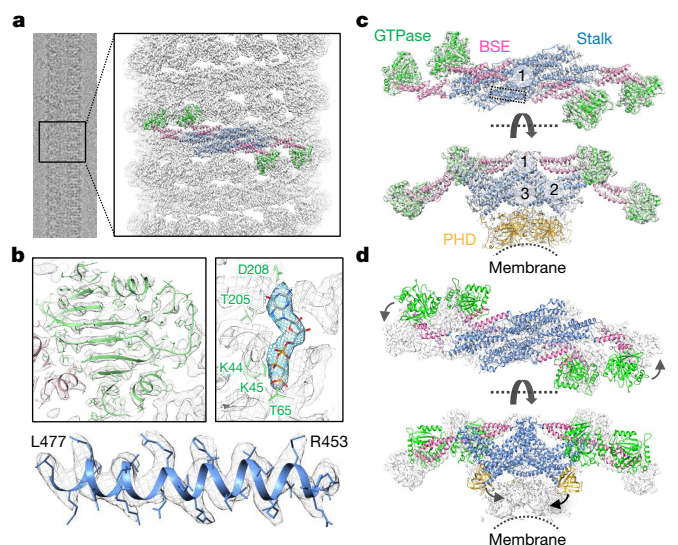


Fig. 1 | Cryo-EM map of assembled dynamin in the GTP-bound state (GMPPCP) on membrane at 3.75 Å. a, Cryo-EM images (left) of helical dynamin tubes were processed to generate a 3D map (right) and subsequently a model of the tetramer was built (green, GTPase domain; pink, BSE; blue, stalk; gold, PHD) (Electron Microscopy Data Bank (EMDB) code: EMD-7957; RCSB Protein Data Bank (PDB) ID: 6DLU). $n = 3$ independent experiments with similar results. **b**, Regions in map showing high-resolution features: β -sheet in the GTPase domain (top left), GMPPCP molecule (top right), and side chains of the L477-R453 helix in the stalk (bottom; dashed box in c). **c**, Tetramer model of assembled dynamin with surrounding density and domains coloured as described above. The assembly interfaces are labelled 1–3. **d**, Comparison of the crystal structure of dynamin in the apo state (coloured as above) with our 3D map (grey).

¹Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD, USA. ²Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA. ³Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA. ⁴Deceased: Shunming Fang. *e-mail: jennyh@nidk.nih.gov

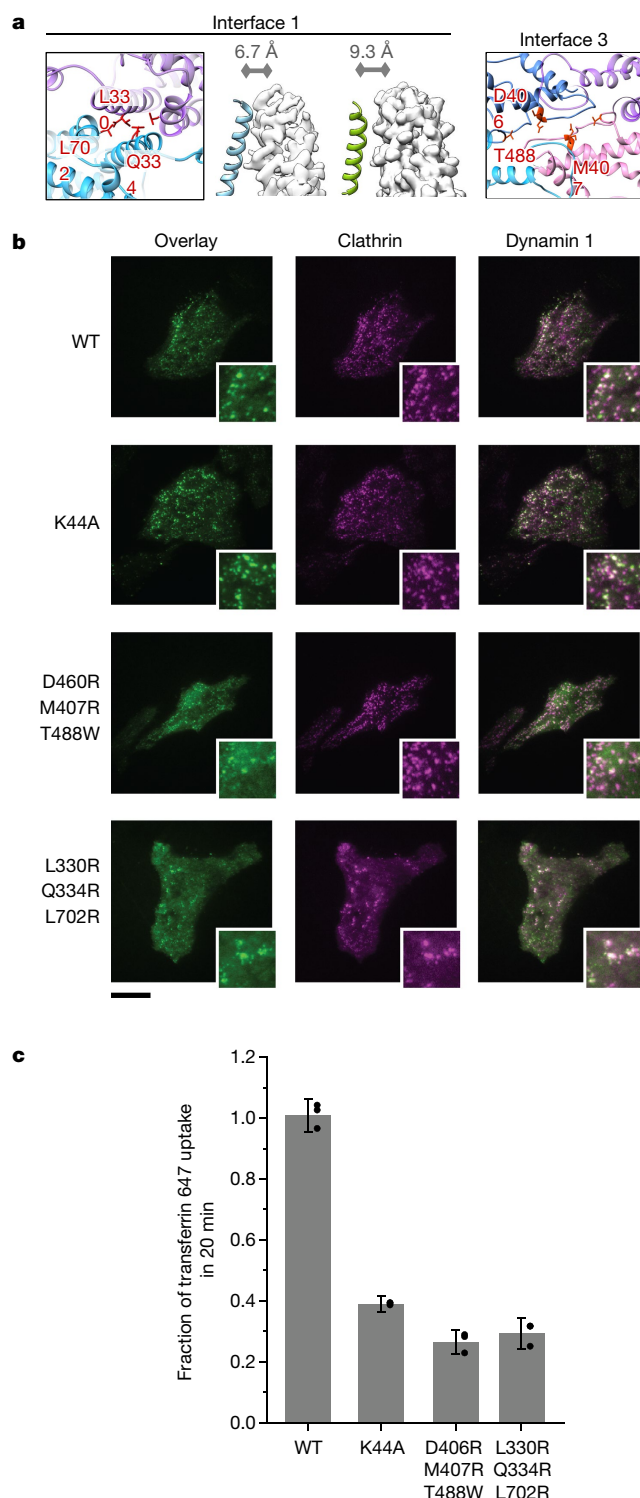


Fig. 2 | Mutations in interfaces 1 and 3 inhibit endocytosis. **a**, Mutations in interface 1 (L330R/Q334R/L702R) and interface 3 (D406R/M407R/T488W) generated for endocytic assays. Middle panels, the $\text{dyn}^{\text{GMPPCP}}$ polymer has a tighter interface 1 (left, blue) than the soluble crystal tetramer (right, green). Distances between stalks in interface 1 are shown above. **b**, Total internal reflectance fluorescence (TIRF) images of dynamin and clathrin colocalization at the plasma membrane ($n = 2$ independent experiments). Scale bar, 20 μm . Insets, 10- μm squares. **c**, Transferrin uptake is defective with interface 1 (L330R/Q334R/L702R) and interface 3 (D406R/M407R/T488W) mutations. Uptake in wild type and K44A mutant are shown for comparison. Mean \pm propagated s.d. from $n = 3$ biological replicates are shown with single replicates (grey dots) background subtracted and referenced to mean values. Trends were verified with $n = 2$ biologically independent experiments (Extended Data Fig. 5).

Each nucleotide treatment of dynamin yielded distinct distributions of polymer diameters (Extended Data Figs. 1a, 2). Consistent with previous reports³, the $\text{dyn}^{\text{GMPPCP}}$ reconstruction has a 40-nm outer diameter and a 7.4-nm inner lumen diameter (Extended Data Fig. 1b). The more constricted dyn^{GTP} reconstruction has a 36-nm outer diameter and a 3.4-nm inner lumen diameter, which is narrow enough to induce spontaneous fission without a protein scaffold¹⁵. In addition, dyn^{GTP} has two-start helical symmetry (Extended Data Fig. 3), similar to a previously published structure of GTP-bound dynamin containing the GTP hydrolysis-deficient mutation K44A¹².

Molecular details of the 3.75 Å map of $\text{dyn}^{\text{GMPPCP}}$ could not be resolved when relying on previously published helical parameters³. New helical parameters (rise 6.3 Å, twist 23.7°) were determined that led to the elucidation of secondary structure, side chains and the nucleotide density as appropriate for the nominal resolution (Fig. 1b), enabling us to derive a precise molecular model of the dynamin tetramer across most of the molecule. The PHDs were of lower local resolution (over 7.0 Å), suggesting they exhibit conformational flexibility and may not conform to a fixed helical symmetry (Extended Data Fig. 1b). This is likely to be due to their unstable positioning on the dynamic lipid membrane while linked to the stalk by long flexible loops, which are disordered in published X-ray crystal structures^{2,10,11}.

Compared to the crystal structure of the lipid-free tetramer, $\text{dyn}^{\text{GMPPCP}}$ adopts an extended form, with the GTPase domain positioned more distal from the stalk and the PHD placed atop the lipid bilayer instead of tucked beneath interface 3 (Fig. 1c, d). The oligomeric interfaces in the stalk domain are similar between the lipid-free tetramer and $\text{dyn}^{\text{GMPPCP}}$ except at interface 1. Interface 1 was originally postulated from the crystal structures of the dynamin-like GTPase Mx¹⁶ but is not clearly defined in crystal structures of dynamin, consisting of only 190 Å² buried solvent-accessible surface area¹¹. By contrast, interface 1 in the $\text{dyn}^{\text{GMPPCP}}$ structure has 726 Å² of buried solvent accessible-surface area (Fig. 2a). To probe the functional importance of interface 1 in endocytosis, we conducted transferrin uptake assays on cells transfected with interface mutants (Fig. 2b, c). Compared to cells with wild-type dynamin, cells with mutations in interface 1 (L330R/Q334R/L702R) exhibited marked endocytosis defects that were similar to those of cells with the GTPase mutation K44A, which is known to disrupt GTP hydrolysis¹⁷, or with interface-3-disrupting mutations (D406R/M407R/T488W). The defects were associated with poor transferrin uptake that did not affect clathrin colocalization¹⁸ (Fig. 2b). This is consistent with recruitment of dynamin to clathrin before dynamin polymer assembly^{19–21} and suggests that either polymerization or the mechano-enzyme function of dynamin were inhibited. Double or single mutations at interface 1 only partially disrupted endocytosis (Extended Data Figs. 4, 5), suggesting that interface 1 is highly robust.

The most notable differences between $\text{dyn}^{\text{GMPPCP}}$ and the lipid-free crystal structures are in the conformations and dispositions of the GTPase and BSE domains (Fig. 1c, d), which are known to depend on nucleotide state. Previously published studies have sought to capture the different dynamin conformational states associated with the GTPase cycle through crystal structures of a dynamin GTPase–BSE dimer (GG)^{3,14}. While interface G2 in $\text{dyn}^{\text{GMPPCP}}$ is equivalent to the GG interface G2 in crystal structures, with an average root mean squared deviation (r.m.s.d.) of 0.8 Å, the GG crystal structures do not fit well into the cryo-EM density of $\text{dyn}^{\text{GMPPCP}}$ (Fig. 3a), suggesting that the cryo-EM structure represents a different hydrolysis intermediate. Notably, the BSE exhibits marked asymmetry across interface G2 in $\text{dyn}^{\text{GMPPCP}}$ (Fig. 3b). Of the two dynamin molecules that form the interface, only one contains a 35° kink centred on T292 in the hinge region linking the BSE and GTPase domains, between helices $\alpha 5^{\text{G}}$ and $\alpha 2^{\text{B}}$ (Fig. 3a, b; Extended Data Fig. 4b). In all dynamin crystal structures, the $\alpha 5^{\text{G}}$ and $\alpha 2^{\text{B}}$ helices are continuous and form an extended helix (T274–E310) with only a slight bend at T294 (Extended Data Fig. 4b), suggesting that a severe kink at T294 is not energetically favourable. The cryo-EM density of the bent BSE is more disordered than that of the unbent BSE, particularly at residues 20–31 (the N-terminal helix and

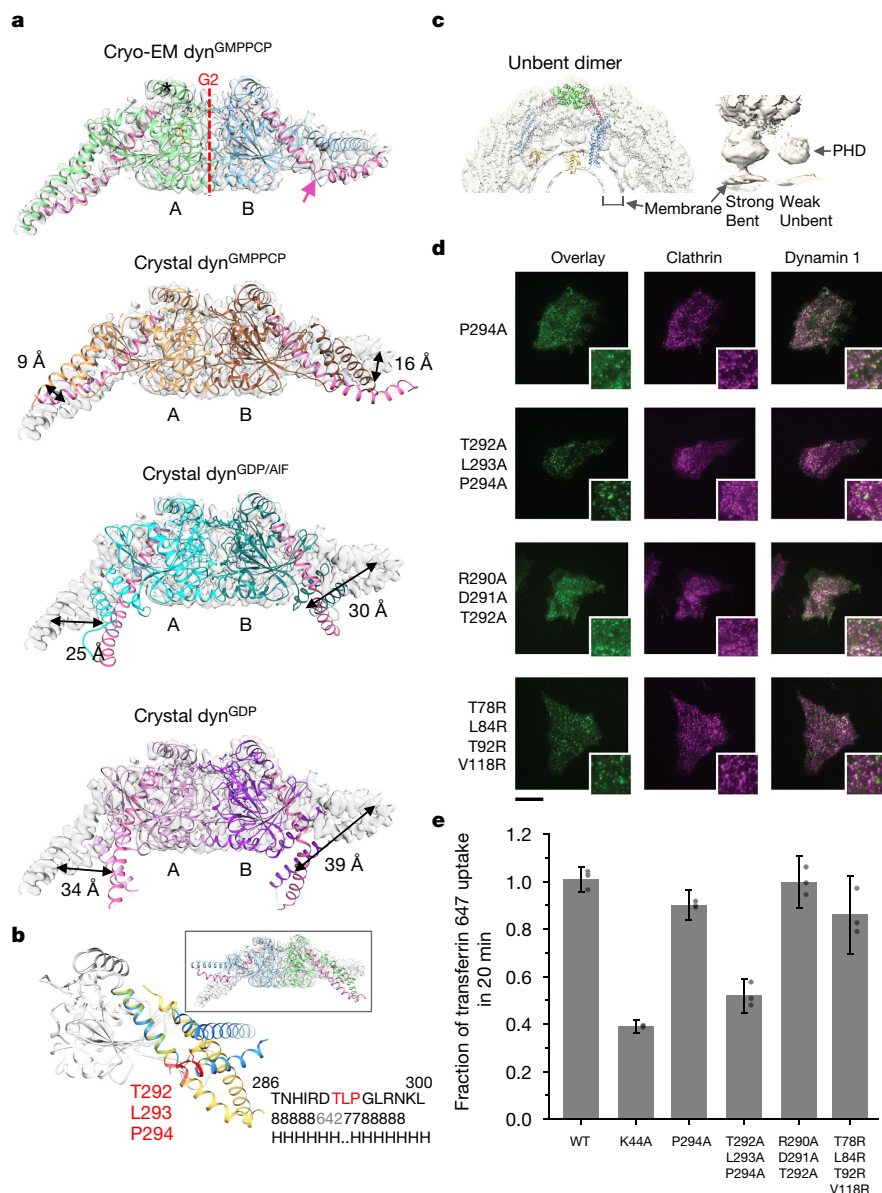


Fig. 3 | Comparison of BSE orientation in relationship to the GTPase domain dimer. **a**, Asymmetry of GG domains in the cryo dyn^{GMPPCP} dimer reveals a unique kink (pink arrow) in the extended helix from the GTPase to the BSE (T274–E310, coloured pink) in the monomer labelled B. Dashed red line indicates interface G2. Comparison of the GG crystal structures in different nucleotide states show a large swing of the BSE and a lack of unique kink in the extended helix in monomer B. From top to bottom: dynamin bound to GMPPCP (PDB ID: 3ZYC), GDP/AIF (PDB ID: 2X2E) and GDP (PDB ID: 5D3Q). **b**, Overlay of cryo-EM dyn^{GMPPCP} GG domains from monomers A and B illustrating the large asymmetry between the BSE domains (coloured blue and yellow for bent and unbent, respectively). Mutated residues T292, L293 and P294 are highlighted in red. Sequence of the helix is shown on the right with helical

loop). To gain additional insight into this, we aligned the coordinates of a dynamin with an unbent BSE to the structure of a bent dynamin in the dyn^{GMPPCP} map at the GTPase domain (Fig. 3c). Unexpectedly, the stalk and PHD of the aligned unbent dynamin were positioned deep inside the lipid bilayer. This result suggests that the BSE bends to accommodate the forces generated at interface G2, which are then transferred across the stalk domain and the PHD to the underlying lipid membrane. Indeed, the cryo-EM density around the PHD of the bent dynamin is better defined than that of the unbent dynamin, as if the PHD from the bent dynamin is stabilized from the transferred force against the lipid membrane (Fig. 3c). To evaluate the functional

propensity calculated by PROFphd (H: α -helix). Insert, flipped GG dimer in map. **c**, Left, a dynamin dimer model with both BSEs unbent (pink). The normally bent dynamin has been replaced with an unbent dynamin aligned at the GTPase domain. Right, the PHDs associated with the bent and unbent BSEs reside in the stronger and weaker densities, respectively. **d**, TIRF images of dynamin and clathrin colocalization at the plasma membrane ($n = 2$ independent experiments). Scale bar, 20 μm . Insets, 10- μm squares. **e**, Transferrin uptake is defective in the helix-stabilizing mutant (T292A/L293A/P294A). Uptake in wild type and K44A mutant are shown for comparison. Mean \pm propagated s.d. from $n = 3$ biological replicates are shown with single replicates (grey dots) background subtracted and referenced to mean values. Trends were verified with $n = 2$ biologically independent experiments (Extended Data Fig. 5).

relevance of bending dynamin, we investigated mutations that disrupt the BSE kink using cell-based endocytosis assays. A triple mutant that increases the helical propensity of the kink (T292A/L293A/P294A)—which presumably resists bending—resulted in substantially reduced transferrin uptake, nearly to the level of the K44A mutation (Fig. 3d, e). The single P294A mutant and the triple R290A/D291A/T292A mutant did not significantly alter transferrin uptake (Fig. 3d, e, Extended Data Figs. 4, 5). An additional mutation on the back side of the GTPase domain (T92R/L84R/V118R/T78R) also had little effect on endocytosis, even though there is close contact here between neighbouring GTPase domains in the assembled polymer (Fig. 3d, e).

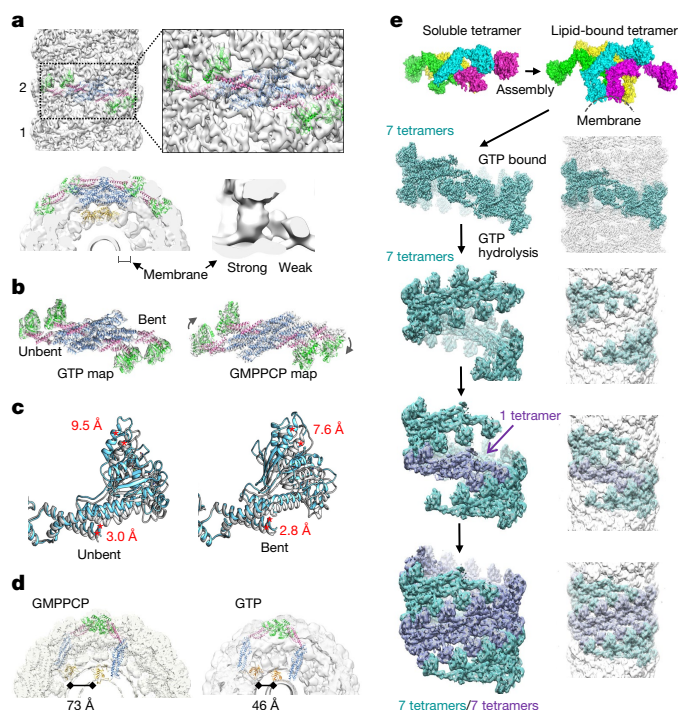


Fig. 4 | 3D map of assembled dyn^{GTP} on membrane at 10.1 Å resolution in the super-constricted state. **a**, In the presence of GTP, dynamin assembles as a two-start helix (labelled 1 and 2). The strong PHD density associates with the bent BSE (EMDB-7958; PDB ID: 6DLV). **b**, Comparison of dyn^{GTP} and $\text{dyn}^{\text{GMPPCP}}$ structures aligned at unbent stalk with the dyn^{GTP} model in dyn^{GTP} density (left) and $\text{dyn}^{\text{GMPPCP}}$ density (right). **c**, Comparison of GTPase–BSE domains from the $\text{dyn}^{\text{GMPPCP}}$ and dyn^{GTP} show an approximately 3 Å shift in the BSE, and 9.6 Å and 7.6 Å movements in the GTPase domains towards the membrane in the unbent and bent monomers, respectively. **d**, A dynamin dimer model made with both BSEs unbent (pink). Comparison between the unbent dimer in the GMPPCP density (left) and the GTP density (right) illustrates potential compression of the dimer upon GTP hydrolysis. The distances between the PHDs are 73 Å and 46 Å for the $\text{dyn}^{\text{GMPPCP}}$ and dyn^{GTP} models, respectively. **e**, Model of dynamin assembly and constriction. The dynamin tetramer unfolds (monomers coloured green, cyan, yellow and magenta) and wraps around the lipid tube in a GTP-bound state as a one-start helix (teal) that is disrupted by GTP hydrolysis, allowing a second strand (purple) to assemble and form a two-start helix.

To understand the role of GTP hydrolysis in the dynamin polymer, a model of a super-constricted dynamin polymer was derived from the 10.1 Å dyn^{GTP} map (Fig. 4a). Whereas the stalk domains of the $\text{dyn}^{\text{GMPPCP}}$ dimer structure fit reasonably well into the dyn^{GTP} density, the GTPase and BSE domains rotate towards the membrane by 10 Å (unbent) and 8 Å (bent) in dyn^{GTP} compared to $\text{dyn}^{\text{GMPPCP}}$ (Fig. 4b, c). Thus, a localized shift in the GTPase and BSE domains induced by GTPase energy mediates super-constriction. An additional effect of this conformational change is apparent when the coordinates of an unbent dynamin are aligned to the structure of a bent dynamin in the dyn^{GTP} map at the GTPase domain (Fig. 4d). Just as for $\text{dyn}^{\text{GMPPCP}}$, the stalk and PHD of the aligned unbent dynamin in the dyn^{GTP} map were positioned deep inside the lipid bilayer but at a much more severe angle (Fig. 4d), suggesting that greater force is being exerted onto the underlying membrane. Furthermore, the underlying lipid bilayer appears to thicken from about 40 Å for $\text{dyn}^{\text{GMPPCP}}$, which matches previous measurements of DOPS (1,2- dioleoyl-sn-glycero-3-phospho-L-serine) lipid bilayers²², to about 46 Å for dyn^{GTP} , which is consistent with greater strain on the lipid²³.

A model of dynamin assembly and constriction emerges from the cryo-EM data (Fig. 4e). In the absence of nucleotide or in the basal hydrolysis state, dynamin polymerizes around lipid tubes, but proceeds to sample a wide range of conformations through structural adjustments

at its interfaces. Previous low-resolution cryo-EM studies²⁴ have shown that interface G2 is formed in this apo state, but probably in a different configuration. Notably, the crystal structure of the apo conformation of dynamin is inconsistent with assembly around a lipid tube (Extended Data Fig. 6). Upon GTP binding, dynamin polymers sample a much more restricted range of conformations, favouring a distinct set of interfaces and a marked asymmetry in the BSE and GTPase domains that applies a force onto the underlying lipid membrane. Localized conformational changes at the GTPase and BSE domains as GTP energy is harnessed drive global changes to the helical symmetry, making room for a second strand to assemble on the membrane tube. This process would require disassembly of dynamin from the lipid bilayer upon GTP hydrolysis, as has been previously reported¹⁴. Furthermore, in crystal structures, interface G2 has been observed only in the presence of GMPPCP or GDP and AlF (aluminium fluoride), but not in the presence of GDP or in the apo state¹⁴. The flexibility of the PHDs should accommodate the transition from the one-start helix to the two-start helix. In summary, these molecular snapshots of the biologically relevant form of dynamin provide a framework for understanding the complex orchestration of GTP-driven conformational changes that mediate membrane constriction.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0378-6>.

Received: 30 June 2017; Accepted: 20 June 2018;

Published online 1 August 2018.

1. Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nat. Rev. Mol. Cell Biol.* **13**, 75–88 (2012).
2. Reubold, T. F. et al. Crystal structure of the dynamin tetramer. *Nature* **525**, 404–408 (2015).
3. Chappie, J. S. et al. A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* **147**, 209–222 (2011).
4. Heymann, J. A. W. & Hinshaw, J. E. Dynamins at a glance. *J. Cell Sci.* **122**, 3427–3431 (2009).
5. Sundborger, A. C. & Hinshaw, J. E. Dynamins and BAR proteins—safeguards against cancer. *Crit. Rev. Oncog.* **20**, 475–484 (2015).
6. Sun, Y. & Tien, P. From endocytosis to membrane fusion: emerging roles of dynamin in virus entry. *Crit. Rev. Microbiol.* **39**, 166–179 (2013).
7. Harper, C. B., Popoff, M. R., McCluskey, A., Robinson, P. J. & Meunier, F. A. Targeting membrane trafficking in infection prophylaxis: dynamin inhibitors. *Trends Cell Biol.* **23**, 90–101 (2013).
8. Daumke, O. & Praefcke, G. J. K. Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Biopolymers* **105**, 580–593 (2016).
9. Sundborger, A. C. & Hinshaw, J. E. Regulating dynamin dynamics during endocytosis. *F1000Prime Rep.* **6**, 85 (2014).
10. Ford, M. G. J., Jenni, S. & Nunnari, J. The crystal structure of dynamin. *Nature* **477**, 561–566 (2011).
11. Faelber, K. et al. Crystal structure of nucleotide-free dynamin. *Nature* **477**, 556–560 (2011).
12. Sundborger, A. C. et al. A dynamin mutant defines a superconstricted pre-fission state. *Cell Reports* **8**, 734–742 (2014).
13. Mattila, J.-P. et al. A hemi-fission intermediate links two mechanistically distinct stages of membrane fission. *Nature* **524**, 109–113 (2015).
14. Antony, B. et al. Membrane fission by dynamin: what we know and what we need to know. *EMBO J.* **35**, 2270–2284 (2016).
15. Kozlovsky, Y. & Kozlov, M. M. Membrane fission: model for intermediate structures. *Biophys. J.* **85**, 85–96 (2003).
16. Gao, S. et al. Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* **35**, 514–525 (2011).
17. Damke, H., Binns, D. D., Ueda, H., Schmid, S. L. & Baba, T. Dynamin GTPase domain mutants block endocytic vesicle formation at morphologically distinct stages. *Mol. Biol. Cell* **12**, 2578–2589 (2001).
18. Larson, B. T., Sochacki, K. A., Kindem, J. M. & Taraska, J. W. Systematic spatial mapping of proteins at exocytic and endocytic structures. *Mol. Biol. Cell* **25**, 2084–2093 (2014).
19. Grassart, A. et al. Actin and dynamin2 dynamics and interplay during clathrin-mediated endocytosis. *J. Cell Biol.* **205**, 721–735 (2014).
20. Srinivasan, S. et al. A noncanonical role for dynamin-1 in regulating early stages of clathrin-mediated endocytosis in non-neuronal cells. *PLoS Biol.* **16**, e2005377 (2018).
21. Sochacki, K. A., Dickey, A. M., Strub, M.-P. & Taraska, J. W. Endocytic proteins are partitioned at the edge of the clathrin lattice in mammalian cells. *Nat. Cell Biol.* **19**, 352–361 (2017).

22. Petrache, H. I. et al. Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **86**, 1574–1586 (2004).
23. Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G. & Engelman, D. M. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Natl Acad. Sci. USA* **101**, 4083–4088 (2004).
24. Chen, Y.-J., Zhang, P., Egelman, E. H. & Hinshaw, J. E. The stalk region of dynamin drives the constriction of dynamin tubes. *Nat. Struct. Mol. Biol.* **11**, 574–575 (2004).

Acknowledgements This work was supported by the NIDDK and NHLBI NIH Intramural Research Program. This work used the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy (NRAMM) located at the New York Structural Biology Center (supported by grants from the Simons Foundation (349247), NYSTAR, and the NIH National Institute of General Medical Sciences (GM103310) with additional support from the Agouron Institute (F00316) and NIH (S10 OD019994-01)), the NHLBI light microscopy core facility, the NHLBI flow cytometry core facility and the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). We thank P. Flicker, J. Kehr, J. Morin-Leisk, A. Shuali and A. Sundborger for discussions; P. Dagur for technical assistance with flow cytometry; and B. Carragher and C. Potter for technical assistance with data collection at NRAMM.

Reviewer information *Nature* thanks S. Scheres and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions L.K. and J.E.H. designed the research; L.K. and J.E.H. prepared protein samples; L.K., H.W., W.J.R. and J.E.H. collected cryo-EM data; L.K., S.F., A.D.K., H.W., B.C. and J.E.H. processed and analysed the data; K.A.S. and J.W.T. designed and performed cell-based assays; M.-P.S. generated all constructs for cell-based assays; L.K. and J.E.H. wrote the paper; all authors were asked to comment on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0378-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0378-6>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.E.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

ΔPRD dynamin expression and purification. HA-tagged ΔPRD dynamin (86 kDa) was expressed in baculovirus-infected TN5 cells and purified as previously described³. In brief, cells were collected after ~48 h and flash frozen in liquid nitrogen. Cell pellets were thawed quickly in ~50 ml of HCB100 (20 mM HEPES, pH 7.2, 100 mM NaCl, 2 mM EGTA, 1 mM MgCl₂, 1 mM DTT) at 37 °C and homogenized by N₂-cavitation at 500 psi for 25 min. The homogenate was diluted with HCB0 (no NaCl) to a final concentration of HCB50 (50 mM NaCl) and then centrifuged for 1 h at 50,000 rpm. To concentrate and enrich for dynamin, 30% ammonium sulfate was added to the supernatant and centrifuged for 12 min at 10,000g. The pellets were resuspended in HCB50, containing protease inhibitors (Roche), and centrifuged at 10,000g for 8 min to pellet aggregated protein. The protein was further purified by a Mono-Q column followed by a Macro-Prep Ceramic Hydroxyapatite (HAP) Type I column. Dynamin was eluted with 400 mM KPO₄ off the HAP column and frozen in liquid nitrogen. The purity was ~95%, judged by Coomassie blue staining, and the final dynamin concentration was 2 mg/ml.

Liposome preparation. Synthetic phosphatidylserine in chloroform (50 μl of 10 mg/ml, DOPS, Avanti) was dried down under argon gas in a glass tube and stored overnight under vacuum to remove excess solvent. The lipid was resuspended in 250 μl HCB150 (150 mM NaCl) and extruded 21 times through a 0.8 μm pore-size polycarbonate membrane (Avanti).

ΔPRD dynamin polymer formation. ΔPRD dynamin polymers were generated as described previously²⁵. Three dynamin treatments were performed to explore a wide range of polymer constriction states (Extended Data Figs. 1, 2). In brief, dynamin was centrifuged at 13,000 rpm (table top centrifuge at 4 °C) for 5 min to remove aggregated protein and then diluted 1:3 with HCB0 for a final concentration ~0.5 mg/ml. The protein was then incubated with DOPS liposomes for 2 h at room temperature with or without 1 mM GMPPCP. For dyn^{GMPPCP} polymers, ΔPRD dynamin was pre-incubated for 5 min before the addition of the DOPS vesicles followed by further incubation for 1 h. For dyn^{GTP} polymers, 1 mM GTP was added to preformed ΔPRD dynamin tubes 5–10 s before freezing.

Cryo-EM sample preparation and imaging. Aliquots of 3.5 μl of each sample were applied to plasma-cleaned (Fishione) C-flat grids (Protochips, CF-1.2/1.3-4C), blotted on the sample side with filter paper for 2 s (22 °C, 90% humidity) and then plunged into liquid ethane with a Leica EM Grid Plunger (Leica Microsystems). For the dyn^{GTP} samples, after 3.5 μl sample was applied to the grids in the grid plunger, GTP was added and plunged into ethane after 5–10 s. The vitrified samples were stored in liquid nitrogen before examination by cryo-EM. For the dyn^{GMPPCP} polymer samples, images were recorded during three sessions on a Titan Krios microscope (FEI) at 300 kV and recorded at 22,500× magnification with a defocus range of 1.0–3.0 μm on a K2 summit camera in counting mode. For the GMPPCP treated sample containing partially constricted polymers and for the dyn^{GTP} sample, images were recorded on a TF20 microscope (FEI) at 200 kV and recorded at 29,000× magnification, with a defocus range of 1.5–3.0 μm on a K2 summit camera in counting mode (Extended Data Table 1).

Cryo-EM data processing. For all images recorded from the FEI Titan Krios microscope, the first frame was removed before motion correction and dose weighting with MotionCor2²⁶. The CTF parameters of non-dose-weighted images were estimated using Ctfind4²⁷, and the correction parameters were applied to the dose-weighted images. For all images recorded from the FEI TF20 microscope, Unblur²⁸ was used for motion correction and dose weighting. The motion-corrected images were then CTF-corrected in Relion using Ctfind4 estimations. From these preprocessed images, well-ordered helical polymers were selected manually in relion 2.0.3 and 2.0.6²⁹ (Extended Data Fig. 2). Polymers adopted a wide range of tubular diameters, and to minimize heterogeneity, all particles were sorted by outer tube diameter, and only particles with the most populated diameter were selected for structure calculation (Extended Data Figs. 1, 2). This was achieved by cross-correlating each particle image with a set of references consisting of helical tubes with varying outer diameters using the Spider software suite^{12,30}. The references consisted of down-sampled images of the dyn^{GMPPCP} from multiple views with varying gaps between the two sides of the polymer. All scripts used for the sorting procedure are available upon request. Relion was used for further particle processing. Dynamin polymers with no nucleotide resisted structural characterization as they were highly disordered and had a wide (41–71 nm) diameter distribution (Extended Data Figs. 1a, 2). The particles were subjected to multiple rounds of 2D classification, with only the highest resolution 2D classes selected after each round. 3D classification did not generate maps with significant differences, possibly owing to the homogeneity achieved from sorting with Spider. For the 3.8 Å dyn^{GMPPCP} map, the B factor automatically estimated in Relion was used (–159.56). For the dyn^{GTP} map, no B factor sharpening was applied. The resolutions of the

final maps were determined using the ‘gold standard’ (FSC = 0.143)³¹ (Extended Data Fig. 1). The mask used to calculate the FSC was determined by choosing the lowest threshold in Chimera³² of one of the unfiltered half-maps that gave no noise outside the reconstruction. Model to cryo-EM map FSC curves were generated by Phenix³³. Helical propensity was calculated by PROFphd³⁴.

3D model refinement. Initial fitting was performed using a model of the dynamin dimer constructed from the crystal structures of the GG-GMPPCP monomer (PDB ID: 3ZYC) and the human dynamin-1 dimer (PDB ID: 3SNH). The model was first docked into the cryo-EM density manually in UCSF Chimera³² followed by rigid-body refinement with Modeller³⁵. Upon convergence, all-atom real space refinement was done using the Phenix-1.13-2998 software suite³³ along with manual model building in Coot 0.8.7³⁶. The final refinement statistics are shown in Extended Data Table 1. Surface burial analysis was performed using the EMBL PISA server³⁷.

Cell culture. All flow cytometry and microscopy was performed on HeLa cells (ATCC CCL-2) that were maintained in phenol red-free DMEM growth media (DMEM, Life Technologies 31053-036; 10% fetal bovine serum, Life Technologies 26140-079; 2 mM glutamax, Life Technologies 35050-061; 1 mM sodium pyruvate, Sigma S8636-100ML) at 37 °C with 5% CO₂. HeLa cells were early passaged stocks directly obtained from ATCC and were tested and shown to be mycoplasma free. The human dynamin1-GFP mutants created for this work were fully sequenced and have been deposited in the Addgene repository database.

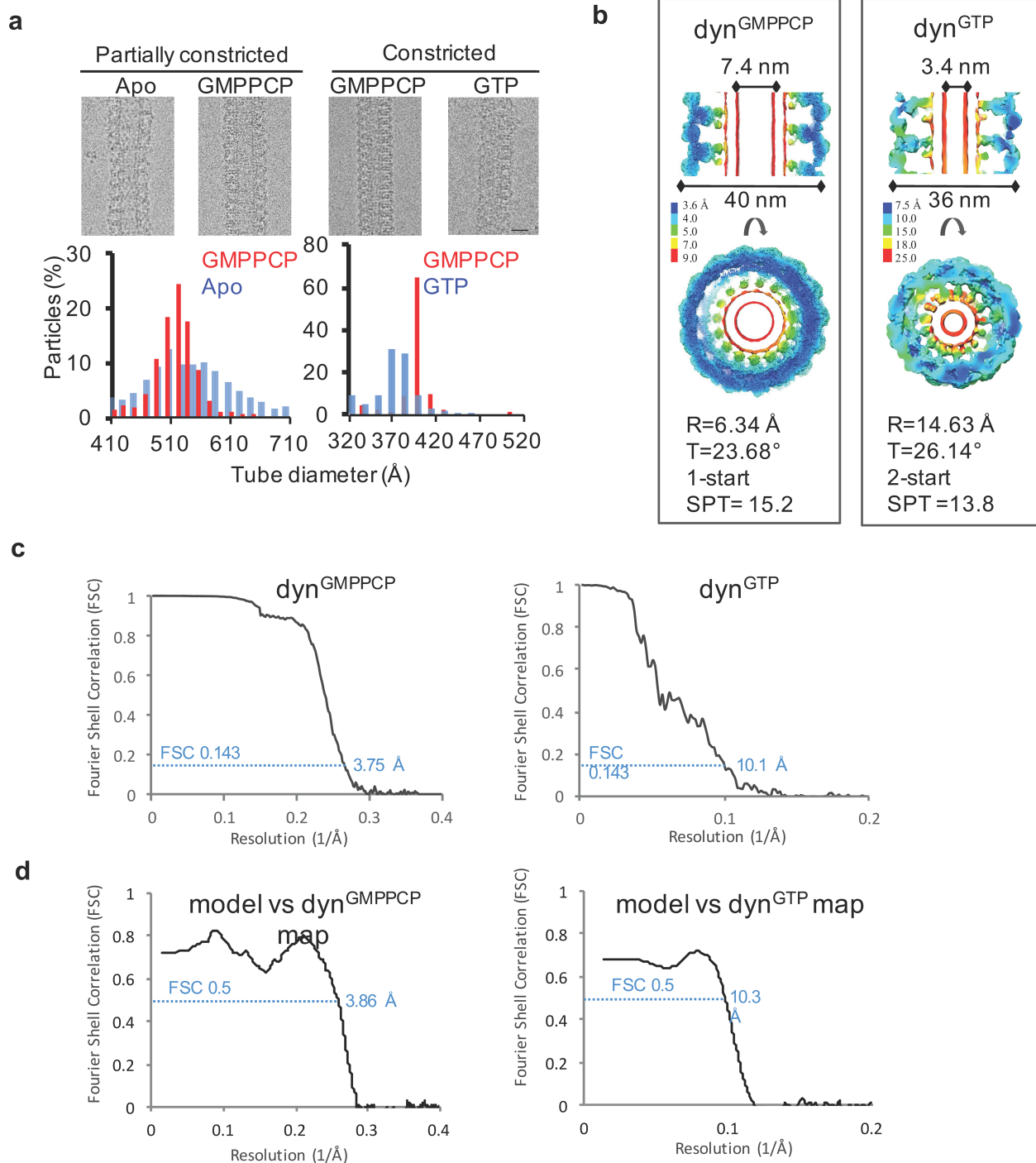
Fluorescence microscopy. Cells were trypsinized (0.25% trypsin, Thermo-Fisher 25200056) and plated on poly-D-lysine-coated coverslips (Neuvitro, GG-25-1.5-pdl) 24–36 h before imaging. Transfection was performed overnight (1–1.5 μg of DNA, 3 μl of lipofectamine 2000 in 0.5 ml of opti-mem and 2 ml of DMEM growth medium). Cells were transfected with the dynamin-GFP mutant of interest and mCherry-clathrin (light chain a, addgene #27680³⁸). TIRF microscopy was performed in imaging buffer (10 mM Hepes, 10 mM glucose, 130 mM NaCl, 2.8 mM KCl, 5 mM CaCl₂, 1 mM MgCl₂, pH 7.4). Cells were imaged on a Nikon Eclipse TI inverted fluorescence microscope with a 100× apoTIRF 1.49 NA objective, 488-nm, and 561-nm excitation lasers. TIRF images displayed in Figs. 2, 3 are characteristic examples from *n* > 20 cells over two independent experiments for each mutant.

Transferrin uptake assay. For the transferrin uptake assay, the cells were prepared as for fluorescence microscopy but were plated onto six-well plates (Fisher Scientific 08-772-1B) and transfected only with the dynamin1-GFP mutant of interest. After overnight transfection, cells were serum starved for 45 min (DMEM; 2 mM glutamax; 1 mM sodium pyruvate). The medium was first replaced with ice cold PBS⁴⁺ (PBS with 1 mM CaCl₂, 1 mM MgCl₂, 0.2% BSA, 5 mM glucose) and placed on ice then replaced by PBS⁴⁺ containing 5 μg/ml Alexa-Fluor 647 conjugated transferrin (Thermo Fisher T23366). The cells were incubated on ice with transferrin for 5 min before the cells were placed into a 37 °C incubator for 20 min. The transferrin was then removed and the cells were rinsed twice with ice cold PBS and incubated on ice with 1 ml, 2 mg/ml pronase (Sigma 10165921001) in PBS for 10 min. The cells were no longer adherent and were pipetted gently to separate clumps before adding 0.25 ml of 16% paraformaldehyde (Electron Microscopy Sciences 15710) to fix for 20 min. The cells were then spun down and resuspended in 300 μl PBS for flow cytometry. Experiments were performed on a BD LSR II flow cytometer and acquired using BD FACSDiva Software version 8.0.1. Single cells were gated away from debris using forward and side scattered light. In one experiment, 1.6 ng/ml DAPI was added to help gate cells away from debris. The results did not differ from a replicate using scattered light (Extended Data Fig. 5). Each experiment when presented together in a plot had identical gating throughout and included each condition in triplicate. The isolated single cells were plotted with Alexa Fluor 647 transferrin fluorescence versus GFP fluorescence. Gating for GFP positive cells and GFP negative cells was chosen based on untransfected controls. Average Alexa Fluor 647 fluorescence was obtained for each population and were background subtracted with a no uptake control. Their ratio determined the fraction of transferrin uptake. In duplicate experiments, the exact uptake ratio could change due to slightly different GFP gating parameters but the relative trends remained constant (Extended Data Fig. 5). In Figs. 2c, 3e, and Extended Data Figs. 4a, 5d, h, the standard deviations are propagated to include the standard deviations from the subtracted background and the GFP-negative reference. The single grey points shown are the average fluorescence from Alexa Fluor 647 transferrin in GFP positive cells in single replicates that have been background subtracted and referenced to the mean values from *n* = 3 replicates.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Data that support the findings of this study have been deposited in EMDB with the accession codes EMDB-7957, EMDB-7958, PDB ID 6DLU and PDB ID 6DLV. The dynamin1-GFP mutant plasmids produced for this study have been deposited at the Addgene plasmid repository.

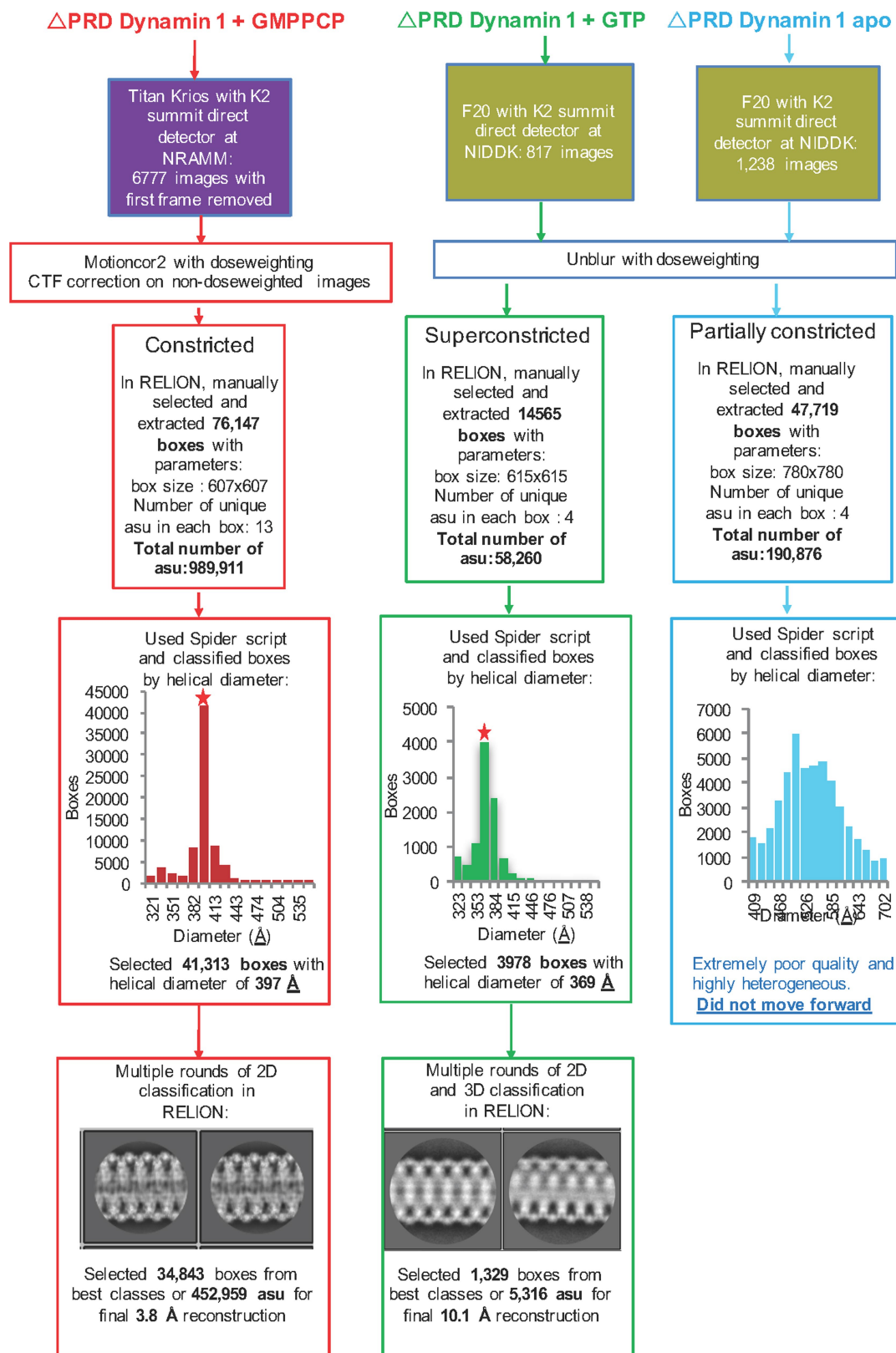
25. Zhang, P. & Hinshaw, J. E. Three-dimensional reconstruction of dynamin in the constricted state. *Nat. Cell Biol.* **3**, 922–926 (2001).
26. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
27. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
28. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
29. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
30. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protocols* **3**, 1941–1974 (2008).
31. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
32. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
34. Rost, B. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266**, 525–539 (1996).
35. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **86**, 2.9.1–2.9.37 (2016).
36. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
37. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
38. Taylor, M. J., Perrais, D. & Merrifield, C. J. A high precision survey of the molecular dynamics of mammalian clathrin-mediated endocytosis. *PLoS Biol.* **9**, e1000604 (2011).



Extended Data Fig. 1 | Cryo-EM parameters and data analysis.

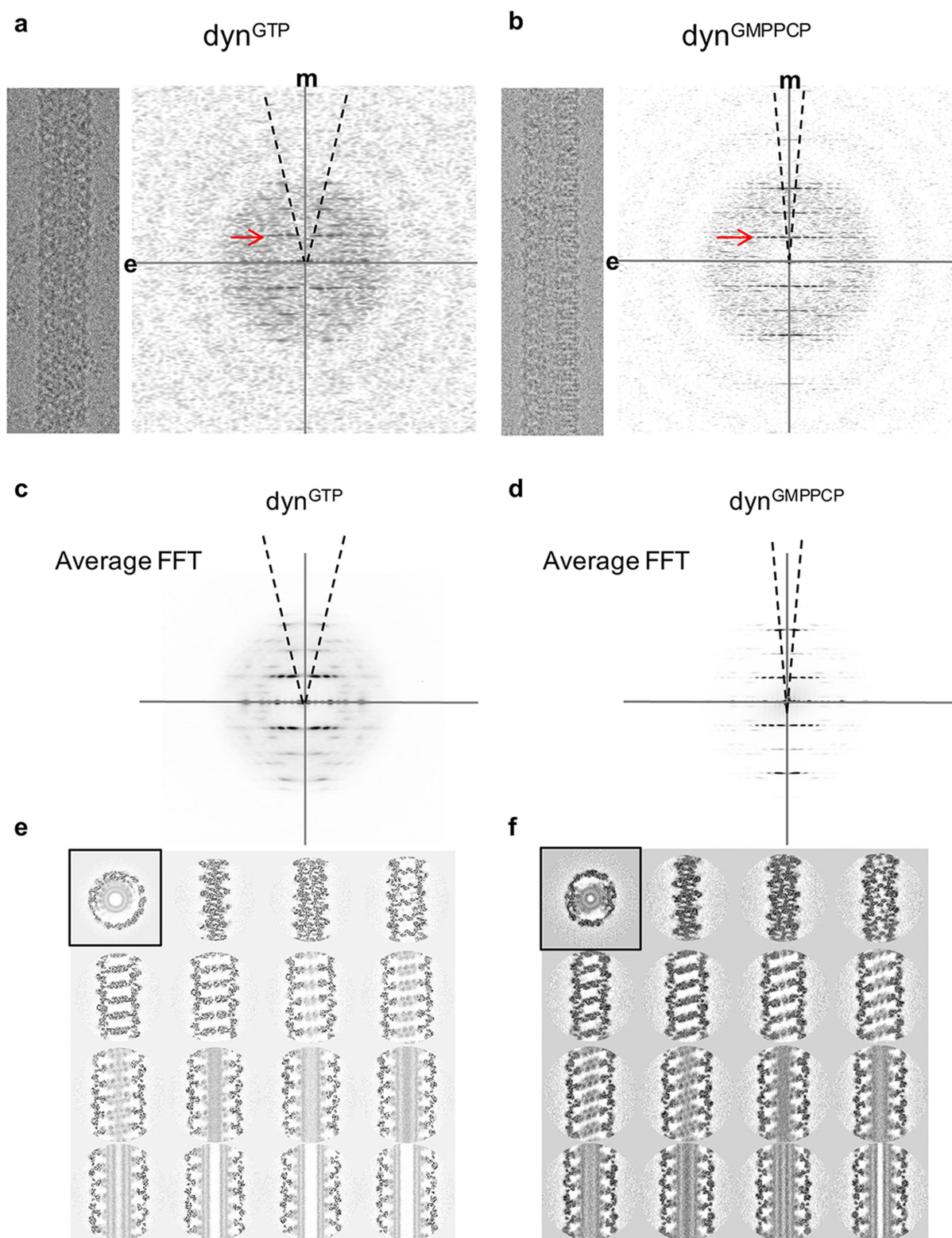
a, Diameter distribution of dynamin tubes in the absence or presence of GMPPCP or GTP. Scale bar, 200 nm. Each experiment was repeated three independent times with similar results. **b**, Local resolution and

helical parameters. R, rise; T, twist; SPT, subunits per helical turn. **c**, Gold standard FSC curves of the dynamin 3D maps. **d**, Model-to-map FSC curves. Dotted blue line indicates gold standard resolution at 0.5 FSC.



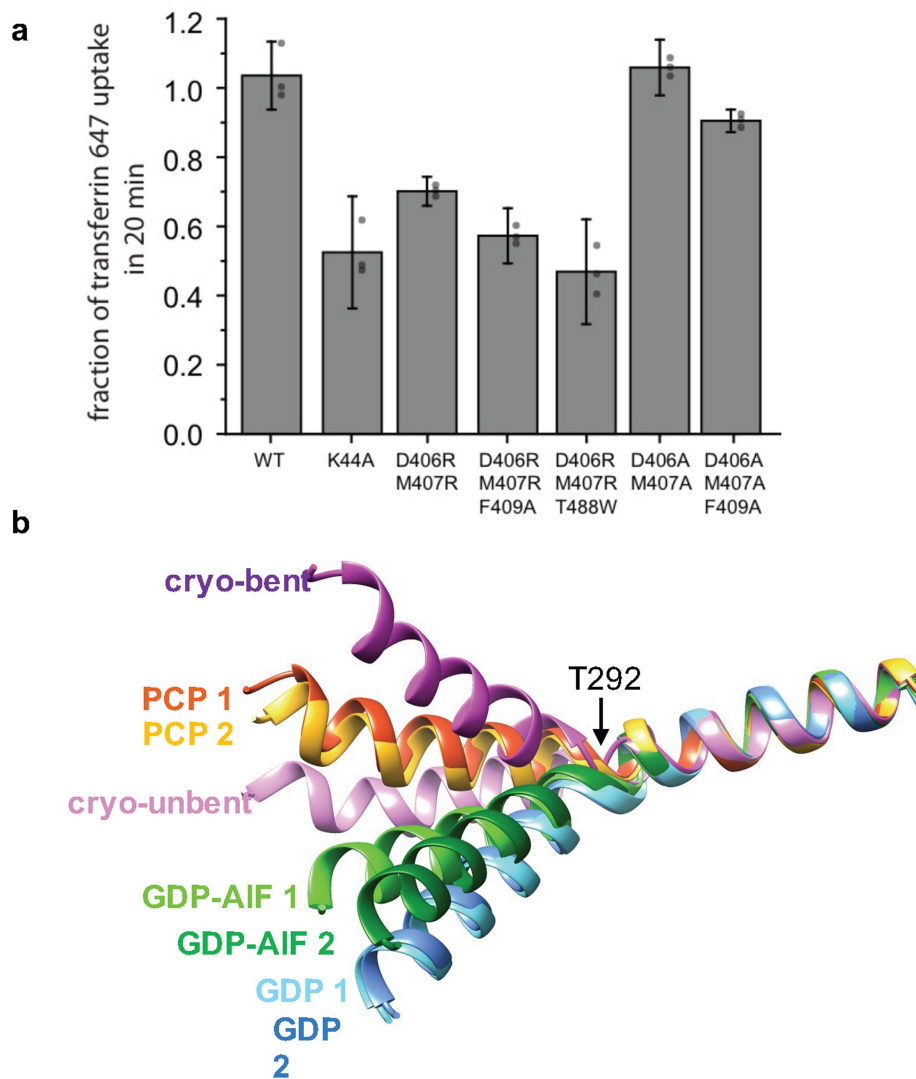
Extended Data Fig. 2 | Cryo-EM data collection and processing flowchart. Starting from the top, the flowchart details the pathways of three separate samples (red, green and blue) of dynamin protein through imaging and processing. The samples were imaged by two different microscopes, and then three different conformational states were selected

manually. Each of these states was processed separately by Spider and then Relion. The red stars in the polymer diameter histograms indicate the diameters of the particles used for the final reconstructions. See Methods for details.



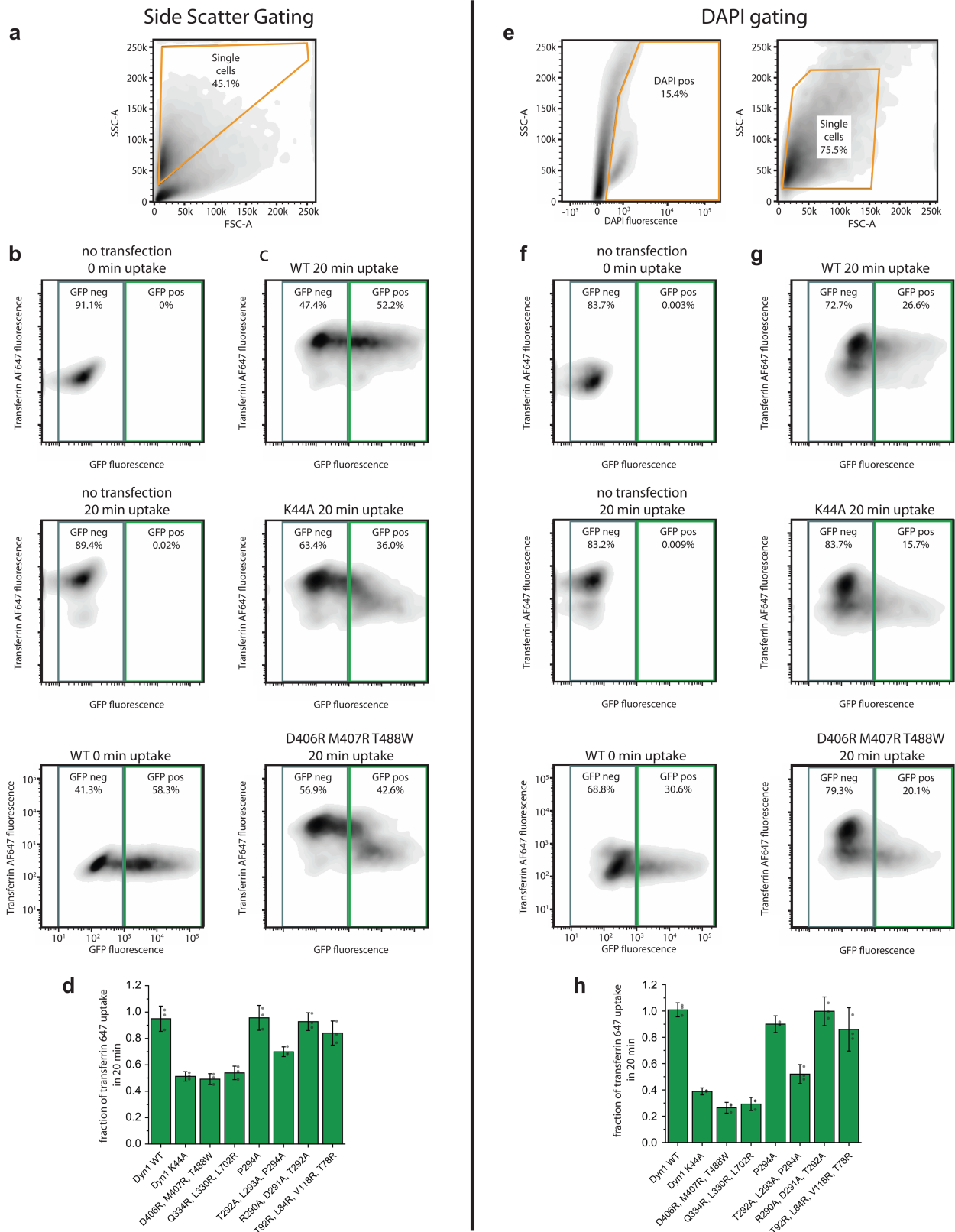
Extended Data Fig. 3 | Dynamin helical tubes and their Fourier transforms. **a, b,** Representative cryo-EM images (left) of two dynamin polymers in different conformations. Each experiment was repeated three independent times with similar results. Right, Fourier transforms of the polymer images. The strong layer lines associated with the pitch (red arrows) are highlighted. The distance of the layer lines from the meridian (m), highlighted by dotted black lines, indicate that the dyn^{GTP} polymer is a two-

start helix, whereas the $\text{dyn}^{\text{GMPPCP}}$ polymer is a one-start helix. **c, d,** Power spectra of 2D class averages from dyn^{GTP} (**c**) and $\text{dyn}^{\text{GMPPCP}}$ (**d**), highlighting the differences between a two-start and one-start helix. **e, f,** Sections of the $\text{dyn}^{\text{GMPPCP}}$ (**e**) and dyn^{GTP} (**f**) maps starting from the outside and going to the middle section. A top middle section looking down the centre of the tubes is shown in the top left panels.



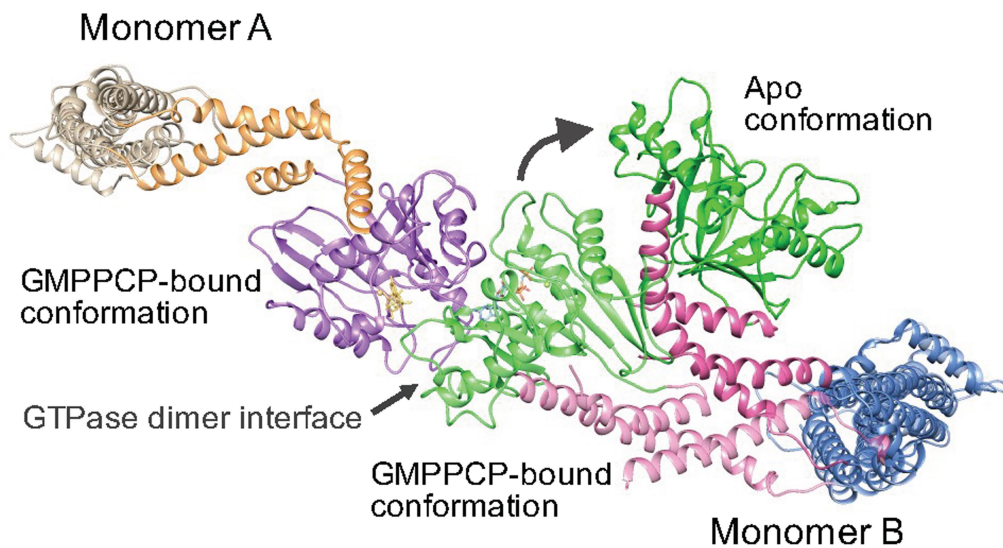
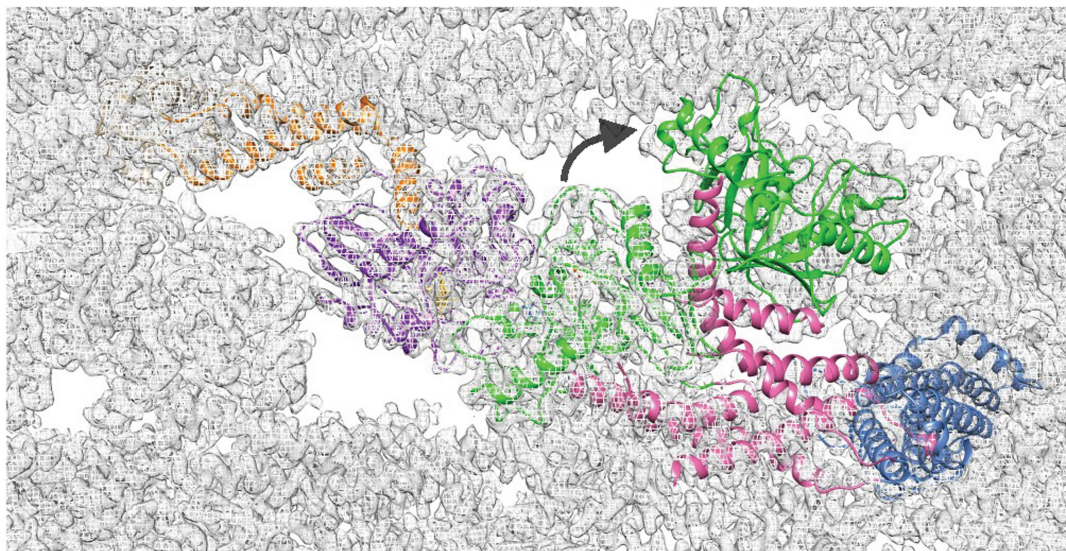
Extended Data Fig. 4 | Functional cell-based assays probing dynamin mutants and structural comparison of the BSE hinge. a, Transferrin uptake of additional interface 3 mutants. Wild-type and K44A are shown for comparison. Mean \pm propagated s.d. from $n = 3$ biological replicates are shown. Also shown are single GFP⁺ biological replicates that are each background subtracted and referenced to mean values (grey dots). Experiment was repeated and trends verified with $n = 2$ biologically

independent experiments. **b,** Comparison of $\alpha 5^G$ and $\alpha 2^B$ helices from dyn^{GMPPCP} (cryo-bent and cryo-unbent, respectively) and available crystal structures, including dynamin bound to GMPPCP (PCP 1 and 2, PDB ID: 3ZYC), dynamin bound to GDP-AIF (GDP-AIF 1 and 2, PDB ID: 2X2E) and dynamin bound to GDP (GDP 1 and 2, PDB ID: 5D3Q). The number after each structure (1 or 2) represents the first or the second member of the dynamin dimer presented in each structure.



Extended Data Fig. 5 | FACS gating. a–h. To test the robustness of our gating, we used either side-scattering (**a–d**) or DAPI (**e–h**) to isolate single cells in two different experimental replicates. In both cases, single cells were isolated from debris (**a, e**). Single cells were then separated into GFP⁻ and GFP⁺ gates (**b, c, f, g**). Controls lacking transfection or transferrin uptake (**b, f**) informed gating choices. Inhibitory mutants exhibited a

characteristic dip in transferrin fluorescence in the GFP⁺ cells (**c, g**). The exact fraction of transferrin uptake was dependent on GFP⁺ gating choices (**d, h**) while qualitative trends were consistent. Mean \pm propagated s.d. from $n = 3$ biological replicates are shown. Also shown are single GFP⁺ biological replicates that are each background subtracted and referenced to mean values (grey dots).

a**b**

Extended Data Fig. 6 | Comparison of dyn^{GMPPCP} and dyn^{APO} at interface G2. **a**, A large swing in the BSE of the tetramer in the apo conformation (PDB ID: 5A3F) disrupts interface G2. The apo tetramer and cryo-EM dyn^{GMPPCP} structure (GMPPCP-bound conformation) were aligned by the stalk. Curved arrow indicates the movement of the

G domain. Domains are coloured green for GTPase, pink for BSE, blue for stalk in monomer B, and purple for GTPase, orange for BSE, tan for stalk in monomer A. **b**, Interface G2, coloured as above, in cryo-EM map of dyn^{GMPPCP} (grey mesh).

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	Dyn ^{GMPPCP} (EMDB-7957) (PDB 6DLU)	Dyn ^{GTP} (EMDB-7958) (PDB 6DLV)
Data collection and processing		
Microscope	FEI Titan Krios	FEI TF20
Magnification	22,500X	29,000X
Voltage (kV)	300	200
Electron exposure (e ⁻ /Å ²)	67	36
Defocus range (μm)	1.0-3.0	1.5-3.0
Pixel size (Å)	1.07	1.27
Image processing software	RELION v2.0.6	RELION v2.0.6
Symmetry imposed	Helical	Helical
Initial particle images (no.) ^a	989,911	58,260
Final particle images (no.) ^a	452,959	14,322
Map resolution (Å)	3.75	10.1
FSC threshold	0.143	0.143
Map resolution range (Å)	3.57-5.67	7.8-21
Helical Parameters		
Inner diameter (nm)	7.4	3.4
Outer diameter (nm)	40.0	36.0
Pitch (Å)	96.4	201.5
Rise (Å)	6.35	14.63
Twist (°)	23.68	26.14
Dynamical dimers per turn (no.)	15.2	13.8
Start (no.)	1	2
Refinement		
Refinement Software	Phenix 1.13-2998	Phenix 1.13-2998
Initial model used (PDB code)	3SNH, 3ZYC	3SNH, 3ZYC
Model resolution (Å)	3.86	
FSC threshold	0.5	
Model resolution range (Å)		
Map sharpening <i>B</i> factor (Å ²)	-146.8	Not used
Phenix Mask CC ^b	0.793	0.789
Model composition	(1 dimer)	(2 dimers)
Non-hydrogen atoms	11,993	22,031
Protein residues	1,453	2,678
Ligand atoms	66	0
<i>B</i> factors (Å ²)		
Protein	88.3	482
Ligands	46.0	N/A
R.m.s. deviations		
Bond lengths (Å)	0.007	0.007
Bond angles (°)	0.829	1.468
Validation		
MolProbity score ^c	1.97	1.76
Clashscore ^c	11.78	9.69
Poor rotamers (%) ^c	0.23	0.57
EM Ringer score ^d	1.92	-0.24
Ramachandran plot		
Favored (%) ^c	94.2	96.3
Allowed (%) ^c	5.8	3.7
Disallowed (%) ^c	0	0

^aNumber of particles is equivalent to number of asymmetric units as calculated by (number of boxes) × (number of unique asymmetric units per box).^bModel-to-map fit (CC_{mask}) as reported by phenix.real_space_refine.^cAs reported by Molprobity (<http://molprobity.biochem.duke.edu>).^dAs reported by Phenix.

Unique features of mammalian mitochondrial translation initiation revealed by cryo-EM

Eva Kummer¹, Marc Leibundgut¹, Oliver Rackham², Richard G. Lee², Daniel Boehringer¹, Aleksandra Filipovska² & Nenad Ban^{1*}

Mitochondria maintain their own specialized protein synthesis machinery, which in mammals is used exclusively for the synthesis of the membrane proteins responsible for oxidative phosphorylation^{1,2}. The initiation of protein synthesis in mitochondria differs substantially from bacterial or cytosolic translation systems. Mitochondrial translation initiation lacks initiation factor 1, which is essential in all other translation systems from bacteria to mammals^{3,4}. Furthermore, only one type of methionyl transfer RNA (tRNA^{Met}) is used for both initiation and elongation^{4,5}, necessitating that the initiation factor specifically recognizes the formylated version of tRNA^{Met} (fMet–tRNA^{Met}). Lastly, most mitochondrial mRNAs do not possess 5' leader sequences to promote mRNA binding to the ribosome². There is currently little mechanistic insight into mammalian mitochondrial translation initiation, and it is not clear how mRNA engagement, initiator-tRNA recruitment and start-codon selection occur. Here we determine the cryo-EM structure of the complete translation initiation complex from mammalian mitochondria at 3.2 Å. We describe the function of an additional domain insertion that is present in the mammalian mitochondrial initiation factor 2 (mtIF2). By closing the decoding centre, this insertion stabilizes the binding of leaderless mRNAs and induces conformational changes in the rRNA nucleotides involved in decoding. We identify unique features of mtIF2 that are required for specific recognition of fMet–tRNA^{Met} and regulation of its GTPase activity. Finally, we observe that the ribosomal tunnel in the initiating ribosome is blocked by insertion of the N-terminal portion of mitochondrial protein mL45, which becomes exposed as the ribosome switches to elongation mode and may have an additional role in targeting of mitochondrial ribosomes to the protein-conducting pore in the inner mitochondrial membrane.

We reconstituted the complete mammalian mitochondrial 55S translation initiation complex from purified porcine mitoribosomal subunits and purified recombinant human mtIF2, naturally occurring leaderless human *MT-CO3* mRNA and aminoacylated and formylated human fMet–tRNA^{Met} stalled with a non-hydrolysable GTP analogue (GTP γ S), and determined its structure at 3.2 Å resolution by cryo-electron microscopy (cryo-EM) (Fig. 1a, Extended Data Fig. 3). Using focused classification, we refined the cryo-EM maps to 3.2 Å and 3.1 Å for the 39S and the 28S subunits, respectively (Extended Data Fig. 2b), which enabled building and refinement of the atomic model (see Methods). The five major domains of mtIF2 are positioned to interact with the decoding centre in the A site of the small ribosomal subunit as well as the sarcin–ricin loop (SRL) and the 3'-CCA end of fMet–tRNA^{Met} close to the peptidyl transferase centre (PTC) of the large ribosomal subunit (Fig. 1a, b). Moreover, we find ribosomal bL12m contacting the mtIF2 GTPase domain (see Extended Data Fig. 3e).

mtIF2 diverges from bacterial IF2 in several functionally important areas, despite having a conserved core fold. The mammalian mitochondrial IF2 contains an insertion of 37 amino acids between domains II and III, which forms an α -helix that extends towards the decoding centre (Fig. 1b). At the decoding centre, the helix kinks and packs

against the A-site mRNA to bridge decoding nucleotides G256 (G530 in *Thermus thermophilus*) and A918/A919 (A1492/A1493 in *T. thermophilus*). This positions a Trp-Lys-X-Arg motif (corresponding to mtIF2 residues 486–489) and an aromatic side chain (residue 494)—both of which are strictly conserved—in front of the A-site codon of the mRNA (Fig. 2a, Extended Data Figs. 4, 5a, b). The mRNA extends into the P site, in which the start codon is bound by the fully accommodated fMet–tRNA^{Met} (Extended Data Fig. 5c). However, there are no specific contacts between the insert and the mRNA in the A site; instead, W486 of mtIF2 stacks on top of rRNA G256, which is retained in a *syn* conformation. The mtIF2 insert then contacts helix h44, causing A919 to flip outwards and to stack with the first base of the mRNA A-site codon, which may prevent the mRNA from sliding. The decoding nucleotide A918 is not flipped outward and resides within an undistorted h44. That h44 remains undistorted during translation initiation in mitochondria contrasts with bacterial and eukaryotic initiation and re-initiation complexes^{6–9}. The interactions of the mtIF2 insert with the A site resemble the interactions of bacterial IF1 with the decoding centre, although mtIF2 adopts a completely different fold⁶ (Fig. 2a). This finding is consistent with an earlier genetic study showing that mtIF2 is able to substitute for *Escherichia coli* IF1 and IF2 in living cells¹⁰, and is in line with a low-resolution structure of mtIF2 on the *E. coli* ribosome¹¹. To clarify whether A-site interaction of the mtIF2 insert is required for the function of the factor, we used a recombinant *E. coli* in vitro expression system¹² (PURE system) that allows substitution of bacterial initiation factors. To ensure proper binding of the mRNA to the bacterial ribosome, the template contained a Shine–Dalgarno sequence. Therefore, the effects we observe are due to processes that occur after mRNA binding to the ribosome. We monitored in vitro translation of the model substrate DHFR–SBP (see Methods, Extended Data Fig. 8a) in reactions lacking *E. coli* IF1 and IF2 but containing mtIF2. Wild-type mtIF2 efficiently replaced bacterial initiation factors (Fig. 2c). Deletion of the Trp-Lys-X-Arg motif strongly diminishes mtIF2 function, suggesting that the mtIF2 insert increases efficiency of translation initiation—probably by excluding elongator tRNAs from premature binding to the A site and by preventing mRNA slippage to ensure correct reading frame selection. These functions are likely to be even more important for the leaderless mRNAs that are present in mitochondria, which do not form stabilizing Shine–Dalgarno interactions with mitoribosomal RNA.

Because mtIF2 does not form specific interactions with mRNA, start codon selection could occur by mitoribosome-specific mRNA engagement and subsequent threading of the mRNA into the mRNA channel for start codon–anticodon interaction. In our 28S cryo-EM map, filtered to lower resolution, *MT-CO3* mRNA is engaged with mitoribosome-specific pentatricopeptide repeat (PPR) protein mS39, which crowns the mRNA entrance (Extended Data Fig. 6a). These contacts may not be sequence- or structure-specific, as 5' sequences of all 11 human mitochondrial mRNAs do not contain a clear consensus sequence and have been shown to exhibit no or only very weak secondary structures¹³. However, starting from codon 7, the mRNAs

¹Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland. ²Harry Perkins Institute of Medical Research, Centre for Medical Research, QEII Medical Centre and School of Molecular Sciences, The University of Western Australia, Nedlands, Western Australia, Australia. *e-mail: ban@mol.biol.ethz.ch

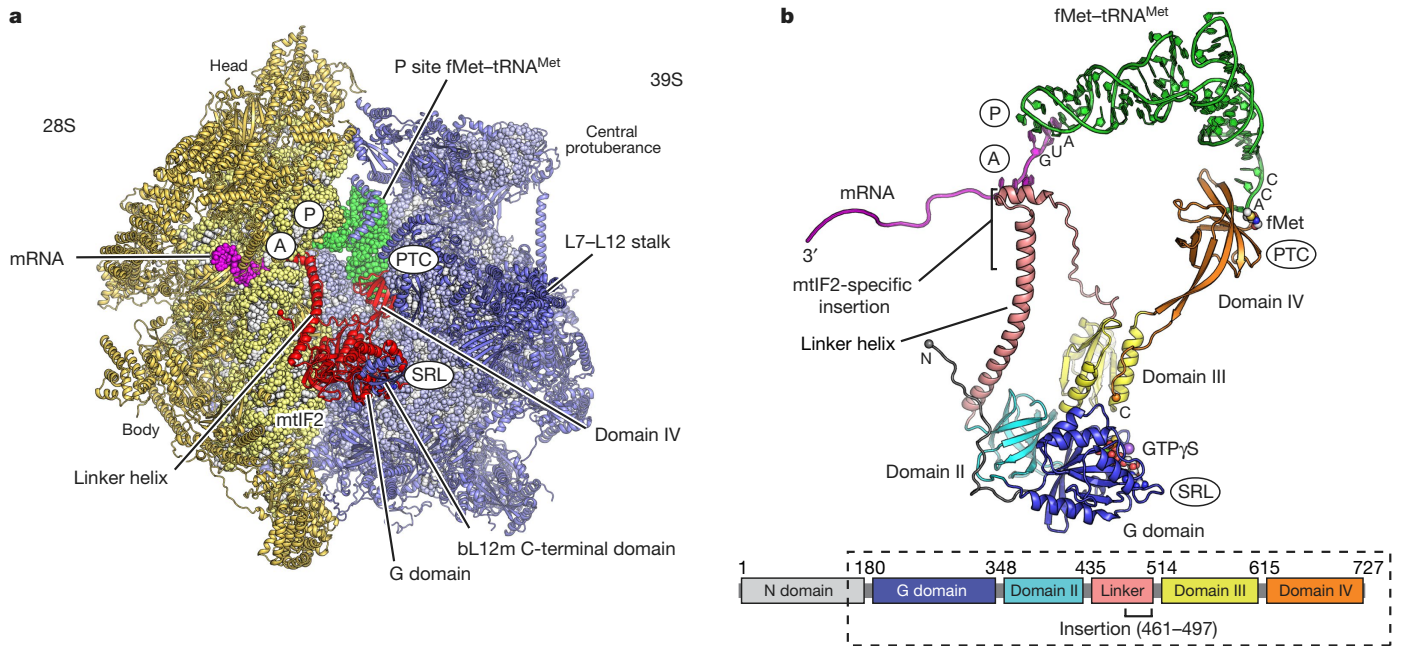


Fig. 1 | Architecture of the mammalian translation initiation complex. **a**, mtIF2 (red) bound between the small ribosomal subunit (28S, yellow) and large subunit (39S, blue) contacting initiator tRNA (green), the sarcin loop (SRL), the peptidyl transferase centre (PTC) and the decoding centre (A and P sites). **b**, Top, the ternary complex (mtIF2, fMet-tRNA^{Met}

and GTP-γS) displayed in isolation, with ribosomal interaction sites indicated. mtIF2 domains are colour-coded according to the schematic representation (bottom). The dashed outline indicates the part of mtIF2 visualized in our structure.

often show U as the second position nucleotide owing to encoding of hydrophobic residues in transmembrane domains (Extended Data Fig. 6d). These U-rich sequences may be the determinant for PPR association and may promote initial binding of the mitochondrial mRNAs

to the initiation complex. The mRNA channel has 'tunnel'-like features and is lined with a series of positively charged conserved amino acids stemming from a mitochondrial-specific extension of uS5m (Fig. 2b, Extended Data Fig. 6b). These interactions of uS5m with the

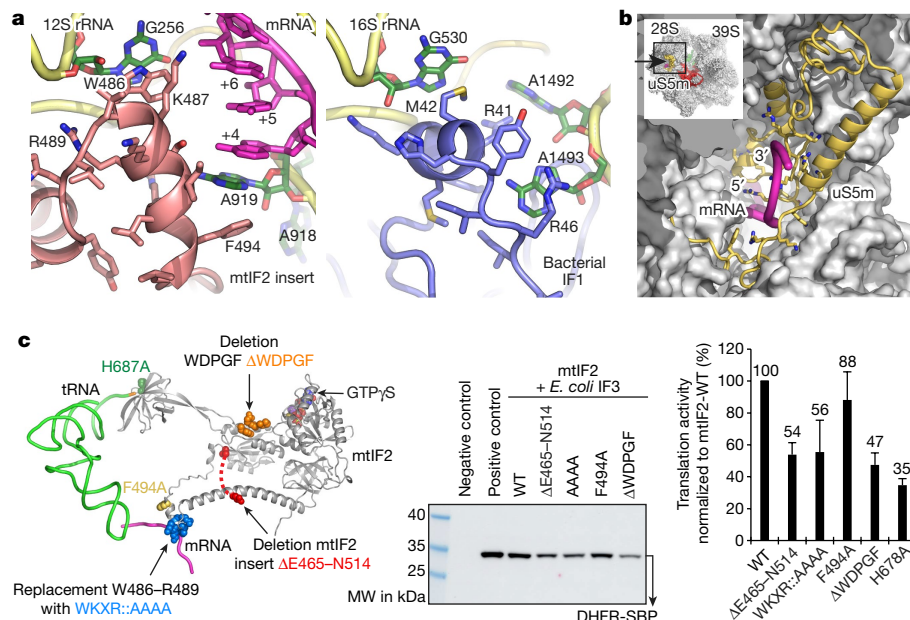


Fig. 2 | Start codon selection of leaderless mitochondrial mRNA. **a**, The mammalian-specific mtIF2 insert (salmon, left) nestles in the A site, where it interacts with decoding bases G256 and A919 in a similar fashion to bacterial IF1 (blue, right; PDB: 1HR0⁶), thereby blocking access to the bound mRNA (magenta). In the mitochondrial complex, A918 resides within helix 44, in contrast to the equivalent residue in the bacterial system, A1492, which flips outwards upon IF1 binding. **b**, The mRNA entry site is surrounded by a mammalian mitochondrial-specific extension of uS5m (yellow), which is rich in positively charged amino acids that guide the mRNA towards the A site. The small inset indicates

the viewpoint. **c**, mtIF2 was substituted in place of bacterial IF1 and IF2 in an in vitro translation assay in *E. coli*. Wild-type (WT) mtIF2 and mutants were compared for efficiency of translation of the model substrate DHFR-SBP. Location and type of mtIF2 mutations are indicated. DHFR-SBP yields after 2 h at 37 °C were quantified after immunoblotting (for mtIF2(H678A), see Extended Data Fig. 8d). The negative control lacks initiation factors whereas the positive control contains *E. coli* IF1, IF2 and IF3. Data are mean ± s.d. of four independent experiments. Yields were normalized to wild-type mtIF2. For gel source data, see Extended Data Fig. 8 and Supplementary Fig. 1.

mRNA via complementary charge and the concomitant narrowing of the mRNA channel identify uS5m as an important component of the mRNA channel positioned between the entrance and the A site. uS5m appears to guide the mRNA towards the P site, where codon–anticodon interaction fixes the AUG and stabilizes the mRNA binding in frame (Extended Data Fig. 5c). Notably, in contrast to the bacterial system¹⁴, and consistent with biochemical data¹⁵, 5' phosphate is not required for recruitment of leaderless mRNA as our mRNA construct loses its 5' phosphate during hammerhead ribozyme cleavage.

In the GTPase domain of mtIF2, switch regions 1 and 2 adopt an ordered conformation and, in conjunction with the P loop, donate residues that form a hexacoordinate arrangement around a Mg^{2+} ion and two water molecules with the β - and γ -phosphates of the bound $GTP\gamma S$ nucleotide (Fig. 3a, Extended Data Fig. 3c). Switch 2 contains the catalytic, highly conserved H238. By analogy with cytosolic ribosomes, interaction with the phosphate backbone of the SRL should orient H238 from its inactive outwards-facing conformation to an active inward-facing conformation to induce GTP hydrolysis¹⁶, even though our maps indicate that H238 can at least partially adopt alternative conformations on the SRL (Fig. 3a). The base of mtIF2 α -helix 12 carries a conserved Y600 residue that was hypothesized to help align the SRL with the GTPase active site of the mtIF2 orthologue in the cytosolic eukaryotic translation initiation complex¹⁷. The side chain of Y600 is oriented towards the catalytic H238, indicating a possible role in facilitating GTP hydrolysis (Fig. 3a). Notably, mtIF2 contains a mitochondrial-specific conserved 723-Trp-Asp-Pro-Gly-Phe-727 motif at its C-terminal tail that is absent in cytosolic orthologues and which directly contacts its switch 2 region, suggesting that the tail influences the position of the catalytic H238 (Fig. 3a, Extended Data Fig. 4). To clarify whether the C-terminal tail is required for mtIF2 function, we tested a mutant lacking the Trp-Asp-Pro-Gly-Phe motif in an in vitro translation assay as described above. Deletion of this mammalian-specific C-terminal Trp-Asp-Pro-Gly-Phe motif leads to a reduction of mtIF2 activity of approximately 50% in the *E. coli* background (Fig. 2c), indicating that it is functionally relevant, presumably by modulating the GTPase activity of the initiation factor.

Mitochondria use only one type of $tRNA^{Met}$, which is used in the form fMet- $tRNA^{Met}$ during initiation and as Met- $tRNA^{Met}$ during elongation. Thus, the sole determinant of aminoacylated $tRNA^{Met}$ that allows mtIF2 to distinguish it from elongator tRNA is the formyl group on the methionine. Formylation of Met- $tRNA^{Met}$ substantially enhances its affinity for mtIF2 and fMet- $tRNA^{Met}$ binding is independent of the nucleotide state of the factor (Fig. 3c). In the structure of the mitochondrial initiation complex, we observe the 3'-CCA end of the $tRNA^{Met}$ charged with formyl-methionine bound to domain IV of the mtIF2 (Fig. 3b). The base of A71 binds into a conserved, mostly hydrophobic pocket and the location and orientation of the methionine side chain can be unambiguously identified with hydrophobic interactions made with the side chains of F632 and A630. In this conformation, addition of the formyl group to the methionine introduces a partial negative charge that is likely to interact tightly with the surrounding D691, H678 and H679. In mtIF2, H678 is universally conserved as a side chain with the capacity to form hydrogen bonds to the formyl group, whereas in the orthologous cytosolic eIF5B, a hydrophobic residue predominates at the equivalent position, consistent with the fact that in the cytosol the methionine of initiator $tRNA^{Met}$ is not formylated and there is therefore no need for specific fMet recognition. Furthermore, in domain II of mtEF-Tu, which is involved in recognition of all amino acids except fMet and shares homology with domain IV of mtIF2, a conserved non-polar amino acid occupies an identical position (Extended Data Fig. 4). Strikingly, we observe that mutation of H678 to alanine abolishes fMet- $tRNA^{Met}$ binding to mtIF2, underlining that stable tRNA binding is critically dependent on specific hydrogen bonding interaction between fMet and mtIF2 (Figs 3c and 2c).

During initiation of translation, the exit of the ribosomal tunnel is generally thought to be vacant, owing to the absence of a nascent chain; however, in our structure, the mitochondria-specific mL45

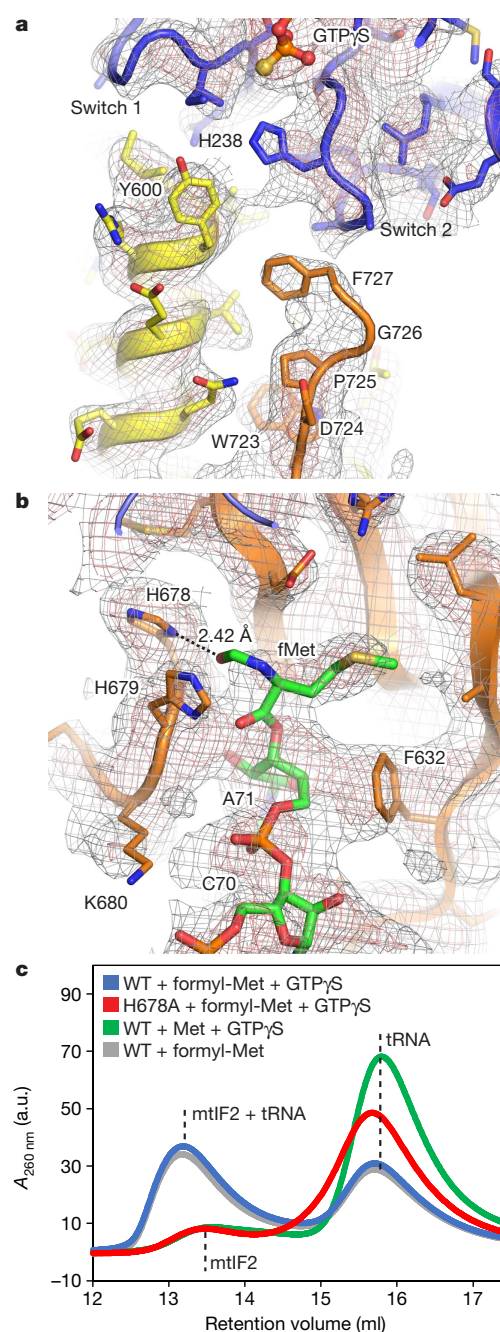


Fig. 3 | mtIF2-specific features regulate its function. a, The C terminus of mtIF2 is extended by a conserved Trp-Asp-Pro-Gly-Phe motif (orange) that reaches towards the catalytic centre of the G domain (blue). Catalytic H238 is shown in a conformation facing the γ -phosphate of $GTP\gamma S$, although our maps indicate that H238 can also adopt an inactive conformation on the ribosome. mtIF2 domain III (yellow) positions the conserved Y600 close to H238 in switch 2 (map at 3 and 6 σ). **b**, Interaction between the $tRNA^{Met}$ CCA-3' end, which carries the formyl methionine (fMet), and domain IV (orange) of mtIF2. H678 stabilizes fMet binding via hydrogen bonding (dashed line). Experimental maps are shown at two contour levels (2 and 3.5 σ). **c**, Size-exclusion chromatography reports on ternary complex formation. $A_{260 nm}$ predominantly detects RNA, indicating that tRNA (23 kDa) runs separately from mtIF2 (72 kDa) if the aminoacylated tRNA is not formylated (green). fMet binding to mtIF2 shifts the tRNA peak to a higher molecular weight (blue). In solution, this interaction occurs independent of $GTP\gamma S$ (grey). Mutation of H678 to alanine abolishes fMet- $tRNA^{Met}$ binding to mtIF2 (red).

inserts its N-terminal tail into the polypeptide tunnel, reaching almost the entire way up to the peptidyl transferase centre (Fig. 4a). The N-terminal tail of mL45, conserved in mammals but absent in

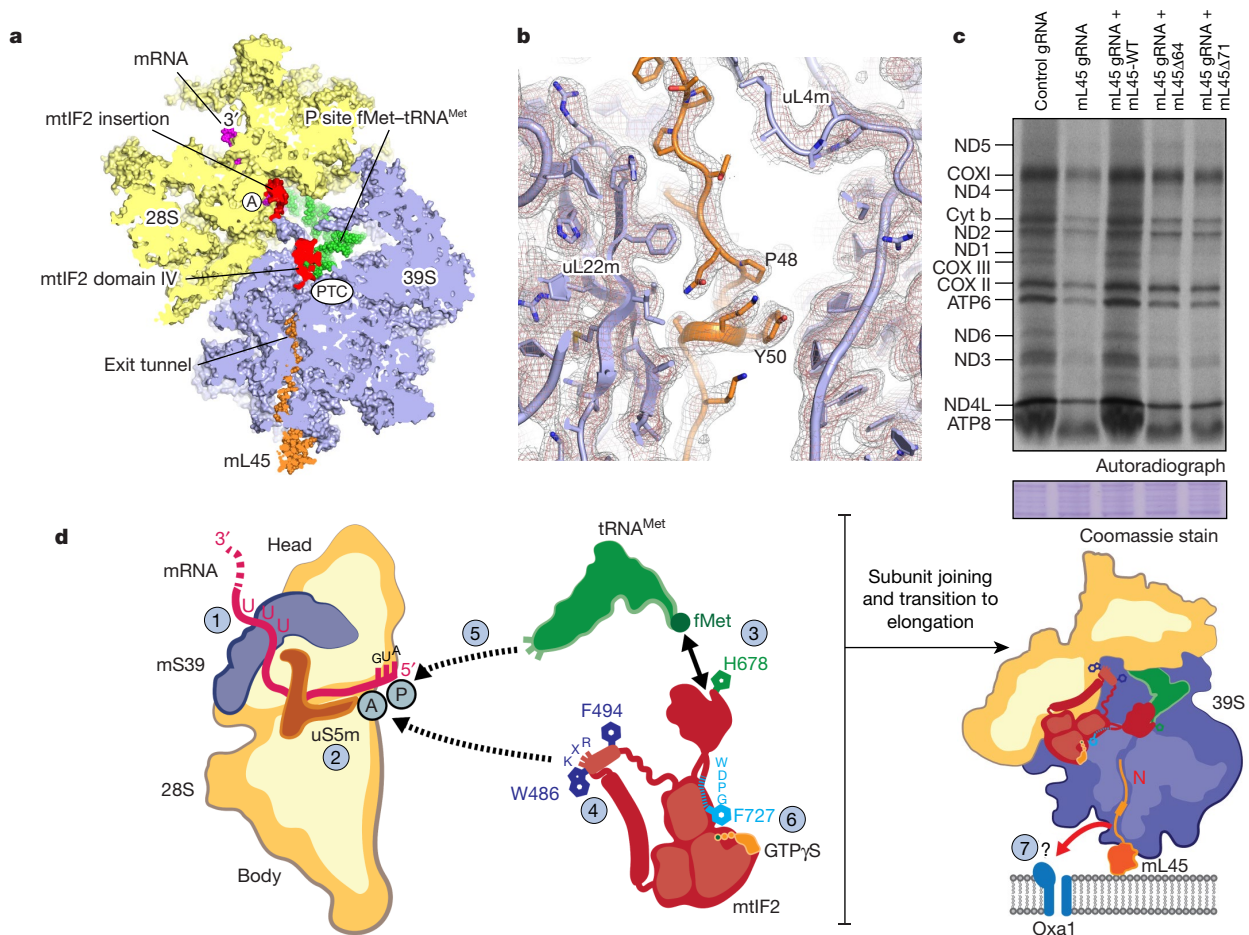


Fig. 4 | Mitochondria-specific mL45 inserts its tail into the exit tunnel.

a, A cutaway view of the 55S translation initiation complex shows that the mL45 (orange) N-terminal extension (NTE) inserts into the vacant exit tunnel. **b**, The mL45 NTE completely blocks the exit tunnel and interacts with the constriction site. Experimental maps at 3 and 7 σ . **c**, Polypeptide synthesis causes displacement of the NTE. The mL45 NTE was truncated at positions G64 and K71 to study its role in vivo. Levels of de novo protein synthesis were measured as described in Methods. Equal amounts of mitochondrial protein (determined by Coomassie staining) were separated by SDS-PAGE and visualized by autoradiography. A representative gel from three independent biological experiments is shown. **d**, Model of mammalian mitochondrial translation initiation. Numbers indicate the steps during complex assembly: 1, association of leaderless mRNA to mitochondria-specific PPR protein mS39; 2, mRNA progression towards

A and P sites assisted by an extension of uS5m; 3, recognition of fMet-tRNA^{Met} by H678 of mtIF2 domain IV; 4, mtIF2 promotes fMet-tRNA^{Met} binding to the small ribosomal subunit and contacts the decoding centre with a mitochondria-specific insertion that shields the mRNA channel and may stabilize mRNA binding; 5, binding of the anticodon of fMet-tRNA^{Met} fixes the reading frame, followed by association to the large ribosomal subunit, facilitated by interactions with the bL12m CTD of the L7-L12 stalk. 39S binding induces GTP hydrolysis in mtIF2 that is; 6, likely to be additionally regulated by a C-terminal extension (F727) of the factor; 7, as the ribosome progresses from initiation to elongation, the N-terminal tail of mL45 has to be displaced and may then form a complex with the insertase Oxa1 to aid insertion and assembly of components of the respiratory chain. For gel source data, see Supplementary Fig. 1.

yeast, encompasses approximately 80 amino acids (from the predicted mitochondrial signal sequence cleavage site at L38 to N115) and is mostly devoid of secondary structure elements (Fig. 4b, Extended Data Fig. 7a, b). The extension contacts proteins uL23m and uL24m at the exit of the tunnel, inserts into the tunnel forming a small helical turn that completely fills the space between uL22m and the 16S rRNA and continues upwards, passing the narrow constriction between proteins uL22m and uL4m with two highly conserved Pro residues (Fig. 4b).

Considering that the N-terminal extension of mL45 completely blocks the tunnel, amino acids K38–N64 must be displaced from the tunnel during the elongation stage of protein synthesis (Extended Data Fig. 7c) and could then fulfill an additional function to promote membrane insertion of nascent chains. To corroborate our hypothesis, we replaced wild-type mL45 with mutants lacking the N-terminal extension in cells. CRISPR–Cas9 deletion of mL45 in HEK293T cells markedly reduced mitochondrial translation, which was recovered with the expression of a wild-type mL45, but not with mL45 lacking amino acids 45–64 or 45–71 from the N-terminal region (Fig. 4c). Levels of

proteins associated with oxidative phosphorylation were also reduced upon mL45 knockout and co-expression of the truncated mL45 proteins, and the levels were rescued only by co-expression of wild-type mL45 (Extended Data Fig. 7d). These results indicate that the N terminus of mL45 is important in mitochondrial translation of membrane proteins. It is possible that the positively charged tail aids recruitment of the translocase Oxa1 to the ribosome, implicating a targeting mechanism analogous to the signal recognition particle, which is essential for the synthesis of membrane proteins in all kingdoms of life, but does not exist in mitochondria^{18–20}.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0373-y>

Received: 27 February 2018; Accepted: 17 May 2018;
Published online 8 August 2018.

1. Greber, B. J. et al. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303–308 (2015).
2. Montoya, J., Ojala, D. & Attardi, G. Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs. *Nature* **290**, 465–470 (1981).
3. Atkinson, G. C. et al. Evolutionary and genetic analyses of mitochondrial translation initiation factors identify the missing mitochondrial IF3 in *S. cerevisiae*. *Nucleic Acids Res.* **40**, 6122–6134 (2012).
4. Christian, B. E. & Spremulli, L. L. Mechanism of protein biosynthesis in mammalian mitochondria. *Biochim. Biophys. Acta* **1819**, 1035–1054 (2012).
5. Kuzmenko, A. et al. Mitochondrial translation initiation machinery: conservation and diversification. *Biochimie* **100**, 132–140 (2014).
6. Carter, A. P. et al. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**, 498–501 (2001).
7. Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leinundgut, M. & Ban, N. The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. *Nat. Struct. Mol. Biol.* **20**, 1015–1017 (2013).
8. Weisser, M. et al. Structural and functional insights into human re-initiation complexes. *Mol. Cell* **67**, 447–456 (2017).
9. Lomakin, I. B. et al. Crystal structure of the human ribosome in complex with DENR-MCT-1. *Cell Reports* **20**, 521–528 (2017).
10. Gaur, R. et al. A single mammalian mitochondrial translation initiation factor functionally replaces two bacterial factors. *Mol. Cell* **29**, 180–190 (2008).
11. Yassin, A. S. et al. Insertion domain within mammalian mitochondrial translation initiation factor 2 serves the role of eubacterial initiation factor 1. *Proc. Natl Acad. Sci. USA* **108**, 3918–3923 (2011).
12. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation system. *Methods* **36**, 299–304 (2005).
13. Jones, C. N., Wilkinson, K. A., Hung, K. T., Weeks, K. M. & Spremulli, L. L. Lack of secondary structure characterizes the 5' ends of mammalian mitochondrial mRNAs. *RNA* **14**, 862–871 (2008).
14. Giliberti, J., O'Donnell, S., Etten, W. J. & Janssen, G. R. A. 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in *Escherichia coli*. *RNA* **18**, 508–518 (2012).
15. Christian, B. E. & Spremulli, L. L. Preferential selection of the 5'-terminal start codon on leaderless mRNAs by mammalian mitochondrial ribosomes. *J. Biol. Chem.* **285**, 28379–28386 (2010).
16. Voorhees, R. M., Schmeing, T. M., Kelley, A. C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835–838 (2010).
17. Fernández, I. S. et al. Molecular architecture of a eukaryotic translation initiation complex. *Science* **342**, (2013).
18. Szyrach, G., Ott, M., Bonnefoy, N., Neupert, W. & Herrmann, J. M. Ribosome binding to the Oxa1 complex facilitates co-translational protein insertion in mitochondria. *EMBO* **22**, 6448–6457 (2003).
19. Keil, M. et al. Oxa1-ribosome complexes coordinate the assembly of cytochrome C oxidase in mitochondria. *J. Biol. Chem.* **287**, 34484–34493 (2012).
20. Denks, K. et al. The signal recognition particle contacts uL23 and scans substrate translation inside the ribosomal tunnel. *New Microbiol.* **2**, (2017).

Acknowledgements We thank T. Schönhut for help with purification of mtIF2 variants; A. Scaiola for support with computational problems and data analysis; V. Godinic for advice on tRNA charging; T. Schönhut, P. Bieri and M. Saurer for help with mitochondria preparation; K. Warinner for help during cloning and for creating a great working atmosphere. We would like to thank the ETH Zürich scientific centre for optical and electron microscopy (ScopeM) for access to electron microscopy equipment and P. Tittmann for technical support. E.K. was supported by an EMBO long-term fellowship (1196-2014). This work was supported by the Swiss National Science Foundation grant (310030B_163478) and via the National Centre of Excellence in RNA and Disease and project funding 138262 to N.B. We thank the NHMRC, ARC and Cancer Council of WA for grants and fellowships to A.F. and O.R. and UWA for a postgraduate scholarship to R.G.L.

Reviewer information Nature thanks L. Sazanov and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions E.K. prepared initiation complex components, ribosomal subunits and cryo-EM samples. D.B. and E.K. acquired the cryo-EM data. E.K. calculated the cryo-EM reconstructions with support from D.B. M.L. built the atomic model and performed coordinate refinement of the atomic model. E.K., M.L. and N.B. interpreted the structure. R.G.L., O.R. and A.F. performed in vivo experiments in HEK cells. E.K. performed in vitro translation assays and size-exclusion chromatography for ternary complex formation. All authors contributed to the final version of the paper.

Competing interests : The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0373-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0373-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to N.B.

Publisher's note : Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Plasmids. Open reading frames for MT-Co3 fused to a hammerhead ribozyme, mtIF2, mtIF3, MetRS and MTF from human mitochondria were ordered from GenScript, codon-optimized for expression in *E. coli* and subcloned into pET24a or pQE-80L vectors, respectively. mtIF2 mutants were generated by site-directed mutagenesis with the exception of mtIF2 (Δ E465–N514), in which amino acids E465–N514 are replaced by the shorter linker from *E. coli* IF2 (E675–H687). mtIF2 (Δ E465–N514) was ordered from GenScript, codon-optimized for expression in *E. coli* and subcloned into pET24a.

Plasmids co-expressing Cas9 and gRNAs were based on pD1311-AD (ATUM), which expresses nuclear localized *Streptococcus pyogenes* Cas9 fused to DasherGFP via the *Thosea asigna* virus 2A peptide. Expression cassettes for gRNAs targeting exon 1 of *mL45* (mL45 gRNA: 5'-ACAAGAGAACCCTTGAGGTA-3') and a control gRNA targeting *EMX1* (EMX1 gRNA: 5'-TGAAGGTGTGGTTCAGAAC-3')²¹ were synthesized from overlapping oligonucleotides and cloned into pD1311-AD. mL45 expression vectors were based on pCI-neo (Promega). The human mL45 ORF (UniProtKB - Q9BRJ2) was subcloned via NheI and NotI restriction sites and silent codon changes were introduced into the gRNA target site to preserve the encoded protein sequence while eliminating the gRNA target site. Truncation mutants of mL45 were made by deleting amino acids I45–G64 or I45–K71 (GenScript) but retaining N-terminal amino acids 1–44, which are for targeting to mitochondria and subsequent proteolytic cleavage of the mitochondrial targeting sequence.

Preparation of porcine mitochondrial subunits. Porcine mitochondria and ribosomal subunits were purified at 4°C as described²², with some modifications. In brief, mitochondria were dissolved in monosome buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 40 mM MgCl₂, 1 mM DTT) and lysed using a Dounce homogenizer. Triton X-100 buffer (monosome buffer including 9.6% (v/v) Triton-X100) was added to a final concentration of 1.6% (v/v) Triton X-100. The suspension was supplemented with 0.5 mM puromycin and stirred for 30 min. The lysate was cleared in two steps at 20,000 r.p.m. using a 45 Ti rotor and subsequently loaded on a 50% (w/v) sucrose cushion. After 24 h at 50,000 r.p.m. (70 Ti rotor) the supernatant was discarded and the ribosome pellet was resuspended in monosome buffer by gently shaking for 1 h on ice. 55S mitoribosomes were further purified on a 10–40% (w/v) sucrose gradient (SW 32 Ti rotor 22,500 r.p.m., 16 h) and fractions containing 55S particles were pooled. Mitoribosomes were concentrated by pelleting (SW 55 Ti rotor, 55,000 r.p.m., 5 h) and resuspended in dissociation buffer (20 mM HEPES-KOH pH 7.6, 300 mM KCl, 5 mM MgCl₂, 1 mM DTT) by gently shaking on ice for 1 h. The suspension was cleared in two steps using a tabletop centrifuge (14,000 r.p.m., 10 min) and loaded onto a 10–40% sucrose gradient (SW 32.1 Ti, 28,000 r.p.m., 14 h). Fractions containing 39S and 28S subunits were pooled separately and subunits were concentrated by pelleting (SW 55 Ti, 55,000 r.p.m., 5 h). Pellets were resuspended in 25 μ l of monosome buffer containing 50 μ M spermine. Rotors and centrifuges were from Beckman Coulter or Eppendorf.

Preparation of human mitochondrial initiation factors 2 and 3. Initiation factors were expressed from a pET24a vector with mtIF2 carrying an N-terminal His₆ tag followed by a TEV cleavage site and mtIF3 carrying a C-terminal His₆ tag preceded by a TEV cleavage site. Proteins were expressed in *E. coli* BL21 SI pRARE at 30°C for 4 h and purified using a HisTrap FF 5-ml column (GE Healthcare) coupled to a HiTrap Heparin HP 5-ml column (GE Healthcare) in standard buffers (50 mM HEPES-KOH pH 7.6, 800 or 50 mM KCl, 5 mM MgCl₂, 10% (w/v) glycerol, 1 mM TCEP, 40 or 500 mM imidazole). The proteins were incubated with TEV protease at 4°C overnight and the His₆ tag, uncleaved initiation factor and TEV protease were removed on a HisTrap HP 1-ml column (GE Healthcare). Proteins were subjected to size-exclusion chromatography on a HiLoad 16/60 Superdex200 (GE Healthcare) and thereby buffer exchanged into storage buffer (40 mM HEPES-KOH pH 7.6, 200 mM KCl, 40 mM MgCl₂, 2 mM DTT, 10% (w/v) glycerol). Initiation factors were then concentrated in an Amicon Ultra-15 centrifugal filter (30-kDa MW cut-off) and flash frozen until further use.

Preparation of human mitochondrial MT-Co3 mRNA. The MT-Co3 gene, encoded as a hammerhead-Co3 construct²³ in a pUC19 vector and under control of a T7 promoter, was digested with StyI to generate a template that was suitable for T7 run-off transcription. The restriction site was chosen such that transcription would yield an mRNA of approximately 200 nucleotides in length²⁴. The template was purified by phenol-chloroform extraction and subsequent ethanol precipitation. In vitro transcription (40 mM Tris-HCl pH 7.2, 30 mM MgCl₂, 0.01% (v/v) Triton X-100, 5 mM DTT, 1 mM spermidine, 10 mM NTPs) was performed at 37°C for 5 h and transcripts were incubated for 1 h at 60°C to complete hammerhead cleavage. mRNA was separated from the hammerhead ribozyme using preparative urea PAGE (5% polyacrylamide, 1 \times TBE, 6 M urea). The appropriate band was excised from the gel, ground into pieces and mRNA was extracted over night at 4°C by shaking the gel pieces in water. The mRNA was then buffer exchanged in an Amicon Ultra-15 (10-kDa MW cut-off) to remove residual urea and concentrate the mRNA before flash freezing it until further use.

Purification of human MetRS and MTF. Human mitochondrial methionyl-tRNA synthetase (MetRS) was purified without the first 42 amino acids, as they contain the mitochondrial import sequence. The protein was expressed from a pQE-80L vector in *E. coli* BL21 carrying the pG-KJE8 chaperone plasmid (TAKARA Bio) at 18°C over night. First, MetRS was affinity purified using an N-terminal His₆ tag (HisTrap FF 5-ml, GE Healthcare) in 50 mM HEPES-KOH pH 7, 300 mM KCl, 20% (w/v) glycerol, 1 mM TCEP and 40 or 500 mM imidazole. The eluted sample from the affinity column was diluted in low salt buffer and applied to a HiTrap Heparin HP 5-ml column (GE Healthcare) using HEPES-KOH pH 7.0, 50 or 500 mM KCl, 20% (w/v) glycerol and 1 mM TCEP. MetRS fractions were pooled and concentrated in an Amicon Ultra-15 (30-kDa MW cut-off) before flash freezing.

Methionyl-tRNA formyl transferase (MTF) was expressed from a pQE-80L vector in *E. coli* BL21 for 5 h at 30°C. The protein carries an N-terminal His₆ tag and was affinity purified on a HisTrap FF 5-ml column (GE Healthcare) in buffers as described for mitochondrial initiation factors and subsequently buffer exchanged into storage buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 10% (w/v) glycerol, 1 mM DTT). MTF was flash frozen and stored until further use.

Purification, charging and formylation of human mitochondrial tRNA^{Met}. Mitochondrial tRNA^{Met} was produced from a construct described as a hammerhead fusion²³. In brief, BstNI digestion was used to generate a template that would result in a CCA-3' end on the transcription product. In vitro transcription, hammerhead cleavage and tRNA purification were carried out as described for the MT-Co3 mRNA. The tRNA^{Met} was stored in water. To induce folding of the tRNA^{Met}, it was diluted to 0.5 mg/ml in water, heated to 80°C for 5 min, supplemented with MgCl₂ to a final concentration of 10 mM and kept at room temperature for 20 min before storing on ice. Leucovorin (Schircks Laboratories) was converted into 10-formyltetrahydrofolate (10-CHO-THF) as described²⁵, to be used as the formyl donor. Aminoacylation and formylation were performed in aminoacylation buffer (50 mM HEPES-KOH pH 7.6, 100 mM NaCl, 10 mM MgCl₂, 5 mM β -mercaptoethanol). Folded tRNA^{Met} was mixed with 2 mM L-methionine, 5 mM ATP, 400 μ M MetRS and incubated for 40 min at 30°C before adding 300 μ M 10-CHO-THF and 1 μ M MTF and keeping the reaction for another 15 min at 30°C. fMet-tRNA^{Met} was purified by phenol-chloroform extraction and subsequent ethanol precipitation. fMet-tRNA^{Met} pellets were dissolved in monosome buffer. Aliquots were flash frozen and stored at –80°C until further use.

Preparation of translation initiation complex. The initiation complex was assembled starting from ribosomal subunits in analogy to canonical eukaryotic or bacterial translation initiation. First, the ternary complex was formed by incubating 10 μ M mtIF2, 10 μ M fMet-tRNA^{Met} and 4 mM GTP γ S for 4 min at 37°C. Then, 60 nM 28S subunits were mixed with 250 nM MT-Co3 mRNA and 250 nM mtIF3. mtIF3 was included to increase efficiency of initiation complex formation²⁶. After 2 min at 37°C, the ternary complex was added in a 1:20 dilution (that is, final concentrations are 500 nM mtIF2, 500 nM fMet-tRNA^{Met} and 200 μ M GTP γ S). In the 55S initiation complex, the 39S subunit (final concentration 60 nM) was added after 3 more min at 37°C and was allowed to associate the small subunit at 37°C for 3 min before placing the initiation complex on ice for 15 min until grid preparation was started. The sample was applied to Quantifoil R2/2 holey carbon grids (Quantifoil Micro Tools) coated with a thin continuous carbon film. The grids were flash frozen in pure ethane on a Vitrobot (FEI).

Data collection and image processing. Images were collected in movie mode on a FEI Titan Krios cryo-electron microscope equipped with a Falcon III direct electron detector (FEI) at 300 kV with a total dose of 40 e/Å² subdivided into 28 frames in 1.4-s exposure using EPU version 1.9.0.30REL (FEI). Images were recorded at 100,719 \times magnification and a defocus range from –1.2 to –2.4 μ m. Movie frames were aligned, summed and weighted by dose in MOTIONCOR2^{27,28} using 5 \times 5 patches, and CTF estimation and particle selection was done using GCTF²⁹ and BATCHBOXER³⁰. Micrographs that contained large pieces of ice or showed poor particle distribution or carbon wrinkles were removed after visual inspection. After CTF estimation we rigorously excluded micrographs that did not reach a resolution higher than 3.2 Å. Particles were picked and initial 3D classification was performed using the 55S mitoribosome (excluding tRNAs in A and P sites) as a reference¹.

We collected three datasets that were first processed separately. Particle images (4 \times binned) were subjected to initial 2D and 3D classification in RELION2³¹ to isolate the population of reassembled 55S ribosomes before further local 3D classifications was used focusing on mtIF2, the small ribosomal subunit and the A site, as shown in the included classification schemes (Extended Data Fig. 1). Masks applied for focused 3D classification in the early steps of particle sorting focused either on mtIF2 G domain, domain II and domain III (mask I) to remove particles that had no mtIF2 bound, or on the connection between mtIF2 domain IV and fMet-tRNA^{Met} (mask II) to remove particles that did not contain tRNA. After initial sorting, our three datasets were joined. We found that

in different classification approaches different parts of the initiation complex were resolved best during further particle sorting and decided to work with 3 different maps that were refined to high resolution in RELION2 and used for model building (Extended Data Figs. 1, 2).

Map 1 and map 2 were generated aligning full size particle images on the small ribosomal subunit in 3D refinement (gold-standard) using a mask surrounding the entire small subunit (mask IV). Subsequently, during focused 3D classification (mask IV) without further alignment particles were removed that still contained no or only poorly associated small ribosomal subunit and the remaining particles were aligned on the small subunit in another round of 3D refinement (mask IV). The mtIF2 insert that resides in the A site is rather flexible, which necessitated one more round of local 3D classification using a mask surrounding the A site that included the small mtIF2 α -helical element deposited in the A site, part of the mRNA and the codon-anticodon pair (mask III). Three-dimensional classification yielded a particle set that showed clear density for codon-anticodon interaction and the mtIF2 insert in front of the mRNA codon located in the A site. These particles were refined either over the entire particle volume (yielding map 1) or using a mask surrounding the small ribosomal subunit (mask IV, yielding map 2). Despite extensive particle sorting, the mediocre local resolution (Extended Data Fig. 2) of the large ribosomal subunit in map 2 also illustrates that the ribosomal subunits show a substantial degree of rotational freedom in the mitochondrial initiation complex. Map 2 was used for general revision of the small ribosomal subunit as well as model building of the mtIF2 insert and the mitochondria-specific extension of uS5m.

Map 3 was generated by another round of focused 3D classification of $4\times$ -binned particle images with the mask surrounding mtIF2 G domain, domain II and domain III (mask I). Three-dimensional classification resulted in two major classes that differed in the degree of subunit rotation and resolution for mtIF2. The class with excellent density for mtIF2 was refined to high resolution over the entire particle volume using unbinned particle images and showed high local resolution used especially for interpretation of the large ribosomal subunit, mL45 and mtIF2. **Structure building and refinement.** The structures of the small 28S and the large 39S ribosomal subunit were built into the cryo-EM maps that had been calculated using focused classification (maps 2 and 3 in Extended Figs. 1a, 2). For this, structures of the porcine subunits (PDB: 4V19, 4V1A and 5AJ3)^{31,32} were docked as rigid bodies, followed by fitting of individual proteins and rRNA segments. The increased resolution and high quality of the maps allowed a general structural update, which included building of more complete rRNA and protein models, adjustments of protein side chains and nucleotide conformers using manual model rebuilding in O³ and COOT³⁴ (Extended Data Table 1). For the previously not decorated PPR folds at the 28S head (mS39) and body (mS27) and rRNA h44, the homologous human mitochondrial structure served as an additional guide (PDB: 3J9M)³⁵. Although resolved to lower local resolution, secondary structure elements of the L7–L12 stalk proteins and the C-terminal domain of bL12 were readily visible, allowing unambiguous docking of Phyre2³⁶ homology models into these regions (PDB: 1ZAV³⁷ for the L7–L12 stalk and 1CTF³⁸ for the bL12-CTD, respectively) (Supplementary Table 1).

Remaining density representing human mtIF2 was initially interpreted by docking IF2 domains of homologous bacterial high-resolution X-ray structures (PDB: 1G7T³⁹, 4KJZ⁴⁰) and an NMR structure of mouse mtIF2 domain IV (PDB: 2CRV). The model was then manually rebuilt and completed, which included extending the linker helix towards the A site and building of the mitochondrial-specific insertion between domain II and III into the cryo-EM map of the small subunit (map 2), where this segment of mtIF2 was better resolved. Owing to lower local resolution, the linker that connects the insertion from the A site back to domain III (residues 496–513) was modelled as UNK (Supplementary Table 1). For the tRNA, a high-resolution structure of tRNA^{Phe} (PDB: 1EHZ⁴¹) served as starting model. The acceptor arm of the human mitochondrial fMet–tRNA^{Met} was then rebuilt into map 3 encompassing the large subunit, where the CCA-3' end with the attached formyl-methionine moiety was resolved at atomic resolution. The tRNA was completed by building the anticodon stem into the small subunit map 2 (Supplementary Table 1).

Phase-restrained coordinate refinement in PHENIX.REFINE⁴² was performed in reciprocal space against the MLHL target using amplitudes and phases back-calculated from the experimental cryo-EM maps as described³². For this, the cryo-EM maps were masked around the individual subunits together with the tRNA and mtIF2, which were present in both models. Each subunit was then individually refined for 7 cycles including rigid body, individual coordinate and B-factor refinement (Supplementary Table 2). Automatic protein secondary structure and RNA base-pair restraints as well as Ramachandran restraints were applied throughout to stabilize the refinement in areas of weaker density. The weighting of the model geometry versus the experimental data (implemented in PHENIX.REFINE as wxc value) was adjusted such that the models displayed excellent geometry with an optimal fit to the cryo-EM map (Supplementary Table 2). The model was validated by calculating the model versus map FSCs using the FSC = 0.5

criterion, showing that the estimated resolutions coincide well with those obtained for the experimental maps at the FSC = 0.143 criterion (Extended Data Fig. 2b).

For the 55S initiation complex, both subunit models were assembled into the map encompassing the entire ribosome. The complete 55S model was then refined in reciprocal space for an additional 3 cycles with refinement settings as described above, followed by 5 cycles of individual B-factor refinement until convergence. The model was validated against the 55S map in a similar manner as depicted for the individual subunits.

Size-exclusion chromatography of the ternary complex. mtIF2 (15 μ M), 4 μ M mitochondrial tRNA^{Met} (aminoacylated or aminocylated and formylated, respectively) and 2 mM GTP γ S were incubated in monosome buffer for 10 min at 37°C and 5 min on ice before separation on a Superdex 200 10/300 GL column (GE Healthcare). Mitochondrial tRNA^{Met} was traced at 260 nm. The presence of isolated tRNA^{Met} in the tRNA peak and as fMet–tRNA^{Met} in complex with mtIF2 in the complex peak was verified by urea or SDS–PAGE, respectively (data not shown).

In vitro translation assays. PURE Express kits were purchased from New England Biolabs with *E. coli* IF1, IF2 and IF3 delivered in separate vials. mtIF2 variants were generated by site-directed mutagenesis and purified as described above. mRNA encoding human DHFR coupled to streptactin binding protein (SBP) and containing an ribosomal binding site (RBS, Extended Data Fig. 8a) was prepared using run-off transcription as described above. Assays were carried out in reaction volumes of 10 μ l according to manufacturer's instructions with DHFR-mRNA as template at a final concentration of 2 μ M. A negative control lacking all *E. coli* initiation factors as well as a positive control containing *E. coli* IF1, IF2, IF3 was included, showing that our generated mRNA was efficiently translated only if all *E. coli* initiation factors were present. The activity of mtIF2 variants in the PURE Express system was characterized in the presence of *E. coli* IF3. First, reactions were incubated at 37°C for 2 h, keeping the IF3 concentration constant and adding mtIF2(WT) at concentrations ranging from 0.05 μ M to 4 μ M to deduce the concentration at which mtIF2 was present in saturating amounts (Extended Data Fig. 8b). We chose to continue our experiments with mtIF2 at 0.3 μ M concentration, since this reflected the middle of the linear activity range, meaning that if mtIF2 variants show stronger or weaker activity than wild type we should be able to detect these differences. Five microlitres of each reaction was applied to 12% SDS PAGE and subsequently immunoblotted using anti-SBP-HRP antibodies (Santa Cruz, SB19–C4). Immunoblots were stained using the ECL Western Blotting Detection Reagent (Amersham) and recorded on an Amersham Imager 600. Immunoblots were quantified using the gel analysis routine in ImageJ (<http://rsb.info.nih.gov/ij/index.html>) and normalized to the activity of mtIF2(WT).

Cell culture and transfections. Human embryonic kidney (HEK293T) cells were cultured at 37°C in humidified 95% air/5% CO₂ in Dulbecco's modified essential medium (DMEM) (Gibco, Life Technologies) containing glucose (4.5 g/l), L-glutamine (2 mM), 1 mM sodium pyruvate and 50 μ g/ml uridine, fetal bovine serum (FBS) (10%, v/v), penicillin (100 U/ml), and streptomycin sulfate (100 μ g/ml). HEK293T cells were plated at 60% confluence in six-well plates and transfected with mammalian expression plasmids in OptiMEM medium (Invitrogen). One hundred and fifty-eight nanograms per centimetre squared of mL45 plasmid DNA and mL45 CRISPR/Cas9 plasmid DNA were transfected using Eugene HD (Roche). Cell incubations were carried out for up to 72 h following transfection and the cells were sorted for fluorescence using a FACSaria II (BD Biosciences). Sorted cells were allowed to recover until they reached 80% confluence and de novo protein synthesis was measured as described below.

The cells were obtained from ATCC, authenticated by STR profiling and found to be free of mycoplasma.

Translation assay. De novo protein synthesis was analysed as previously described⁴³. In brief, HEK293T cells were grown in six-well plates until 80% confluent and de novo protein synthesis was analysed. The growth medium was replaced with methionine and cysteine-free medium containing 10% dialysed FBS for 30 min before addition of 100 μ g/ml emetine for 5 min. Next, 200 μ Ci Express35S Protein Labelling Mix [35S] (Perkin–Elmer) was added and incubated at 37°C for 1 h, then cells were washed in PBS and collected by trypsinization. The cells were suspended in PBS, 20 μ g of proteins were separated on 12.5% SDS PAGE and the radiolabelled proteins were visualized on film.

Steady-state levels of oxidative phosphorylation complexes. Specific proteins were detected using a rabbit monoclonal antibodies against mL45 (16394-1-AP) and bS16m (16735-1-AP), and mouse monoclonal antibodies: Total OXPHOS Cocktail Antibody (ab110412) and β -actin (ab8226). All primary antibodies were diluted 1:1000 using Odyssey blocking buffer (LI-COR). IR Dye 800CW Goat Anti-Rabbit IgG or IR Dye 680LT Goat Anti-Mouse IgG (LI-COR) secondary antibodies (diluted 1:10000) were used and the immunoblots were visualized using the Odyssey Infrared Imaging System (LI-COR).

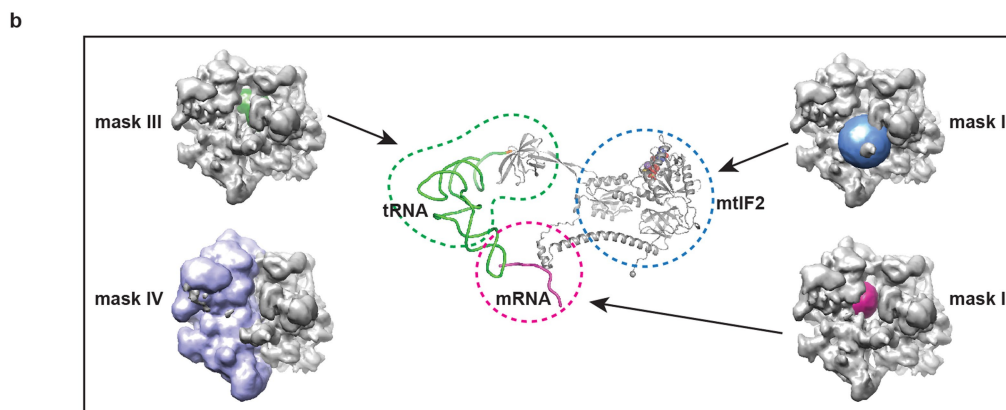
Sucrose gradients of mitochondrial ribosomes to analyse mL45 mutant incorporation. Sucrose gradient fractionation was carried out on mitochondria isolated from cells, as previously described⁴⁴.

Figure generation. Molecular graphics were generated using PyMOL (Schroedinger) or the UCSF Chimera package⁴⁵.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The coordinates and corresponding cryo-EM maps were deposited in the Protein Data Bank (PDB) and in the Electron Microscopy Data Bank (EMDB) under accession codes 6GAZ and EMD-4369 (28S small ribosomal subunit), 6GB2 and EMD-4370 (39S large ribosomal subunit), and 6GAW and EMD-4368 (55S initiation complex). All other data can be obtained from the corresponding authors upon reasonable request.

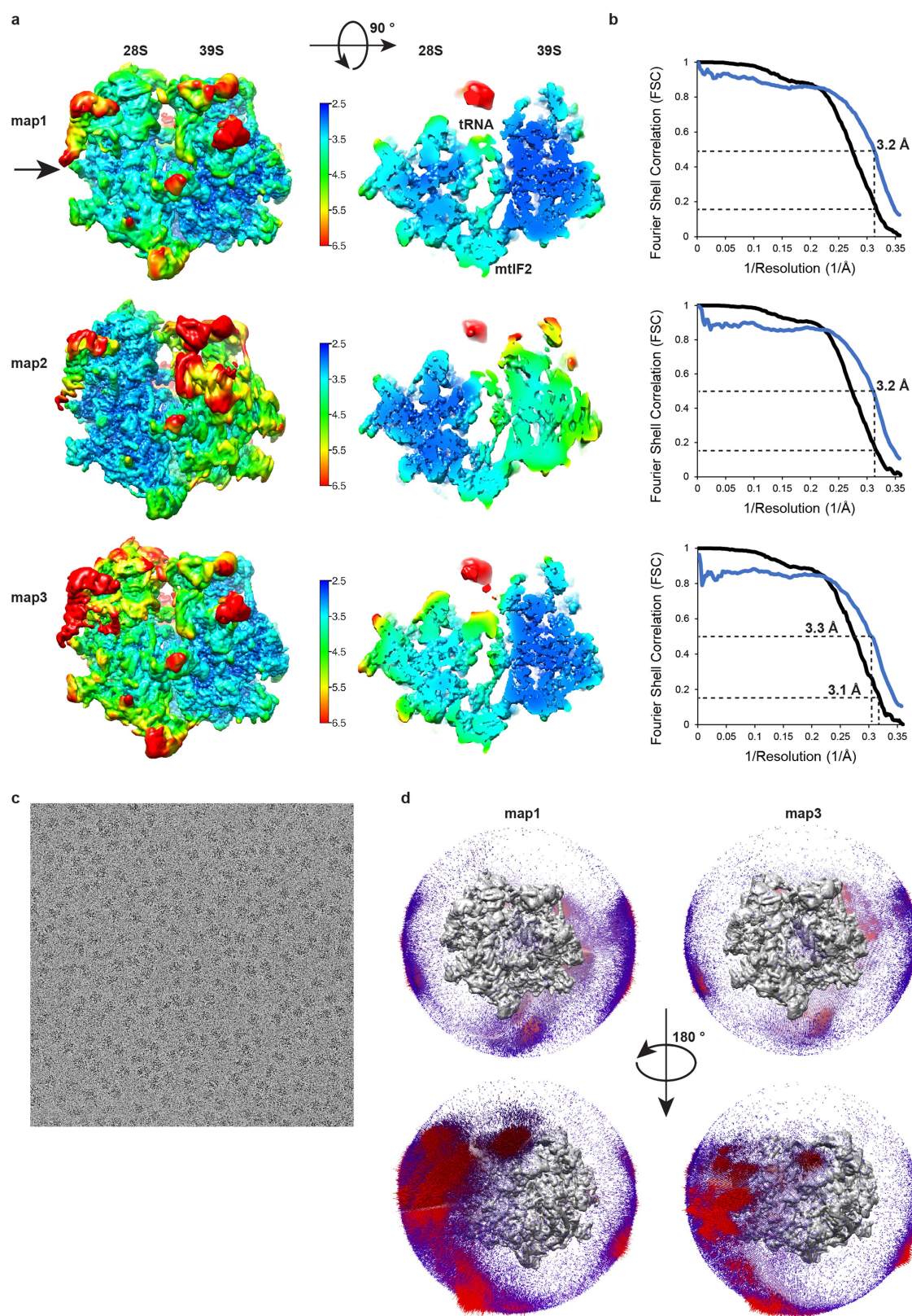
21. Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
22. Greber, B. J. et al. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**, 515–519 (2014).
23. Spencer, A. C., Heck, A., Takeuchi, N., Watanabe, K. & Spremulli, L. L. Characterization of the human mitochondrial methionyl-tRNA synthetase. *Biochemistry* **43**, 9743–9754 (2004).
24. Liao, H. X. & Spremulli, L. L. Effects of length and mRNA secondary structure on the interaction of bovine mitochondrial ribosomes with messenger RNA. *J. Biol. Chem.* **265**, 11761–11765 (1990).
25. Blanquet, S., Dessen, P. & Kahn, D. Properties and specificity of methionyl-tRNA^{Met} formyltransferase from *Escherichia coli*. *Methods Enzymol.* **106**, 141–152 (1984).
26. Koc, E. C. & Spremulli, L. L. Identification of mammalian mitochondrial translation initiation factor 3 and examination of its role in initiation complex formation with natural mRNAs. *J. Biol. Chem.* **277**, 35541–35549 (2002).
27. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, (2015).
28. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
29. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
30. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
31. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
32. Greber, B. J. et al. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283–286 (2014).
33. Jones, T. A. Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr. D* **60**, 2115–2125 (2004).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).
36. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
37. Diaconu, M. et al. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991–1004 (2005).
38. Leijonmarck, M. & Liljas, A. Structure of the C-terminal domain of ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. *J. Mol. Biol.* **195**, 555–579 (1987).
39. Roll-Mecak, A., Cao, C., Dever, T. E. & Burley, S. K. X-ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. *Cell* **103**, 781–792 (2000).
40. Eiler, D., Lin, J., Simonetti, A., Klaholz, B. P. & Steitz, T. A. Initiation factor 2 crystal structure reveals a different domain organization from eukaryotic initiation factor 5B and mechanism among translational GTPases. *Proc. Natl Acad. Sci. USA* **110**, 15662–15667 (2013).
41. Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
42. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
43. Rackham, O. et al. Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells. *Nucleic Acids Res.* **37**, 5859–5867 (2009).
44. Rackham, O. et al. Hierarchical RNA processing is required for mitochondrial ribosome assembly. *Cell Reports* **16**, 1874–1890 (2016).
45. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
47. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D* **73**, 496–502 (2017).
48. Wahl, M. C. & Möller, W. Structure and function of the acidic ribosomal stalk proteins. *Curr. Protein Pept. Sci.* **3**, 93–106 (2002).
49. Helgstrand, M. et al. The ribosomal stalk binds to translation factors IF2, EF-Tu, EF-G and RF3 via a conserved region of the L12 C-terminal domain. *J. Mol. Biol.* **365**, 468–479 (2007).
50. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, (2013).
51. Pallesen, J. et al. Cryo-EM visualization of the ribosome in termination complex with apo-RF3 and RF1. *eLife* **2**, (2013).
52. Kuhle, B. & Ficner, R. A monovalent cation acts as a structural and catalytic cofactor in translational GTPases. *EMBO J.* **33**, 2547–2563 (2014).
53. Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
54. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acid Res.* **32**, W665–W667.
55. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
56. Li, L. et al. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **4**, (2012).
57. Englmeier, R., Pfeffer, S. & Förster, F. Structure of the human mitochondrial ribosome studied *in situ* by cryoelectron tomography. *Structure* **25**, 1574–1581 (2017).



Extended Data Fig. 1 | Classification scheme. a, Schematic representation of how the different cryo-EM maps have been calculated in RELION^{31,47}. Classification yielded three maps used for model building. Maps 1 and 2 were calculated from identical particles with map 1 being refined over the entire particle volume whereas map 2 refinement was focused on the 28S subunit of the ribosome. Map 3 derived from a different particle subset and was refined over the entire particle volume. Resolutions have been estimated in RELION by post-processing the entire

ribosome (map 1), 28S including mtIF2 and tRNA (map 2, indicated in red) or 39S including mtIF2 and tRNA (map 3, indicated in red).

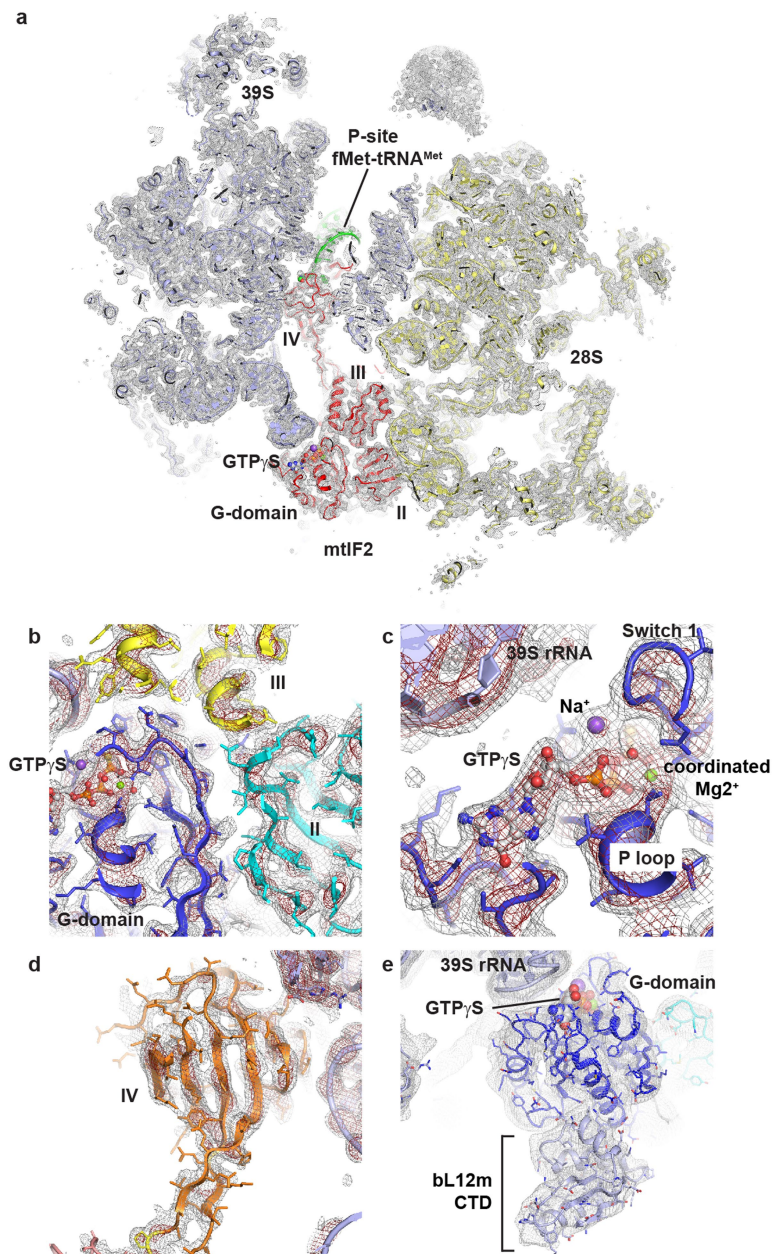
b, Depiction of masks that have been applied for focused 3D classification or 3D refinement in RELION. Mask I encompasses the mtIF2 G domain, domain II and domain III. Mask II includes the 28S A site and the mtIF2 insert. Mask III focuses on fMet-tRNA^{Met} and domain IV of mtIF2. Mask IV includes the 28S subunit.



Extended Data Fig. 2 | See next page for caption.

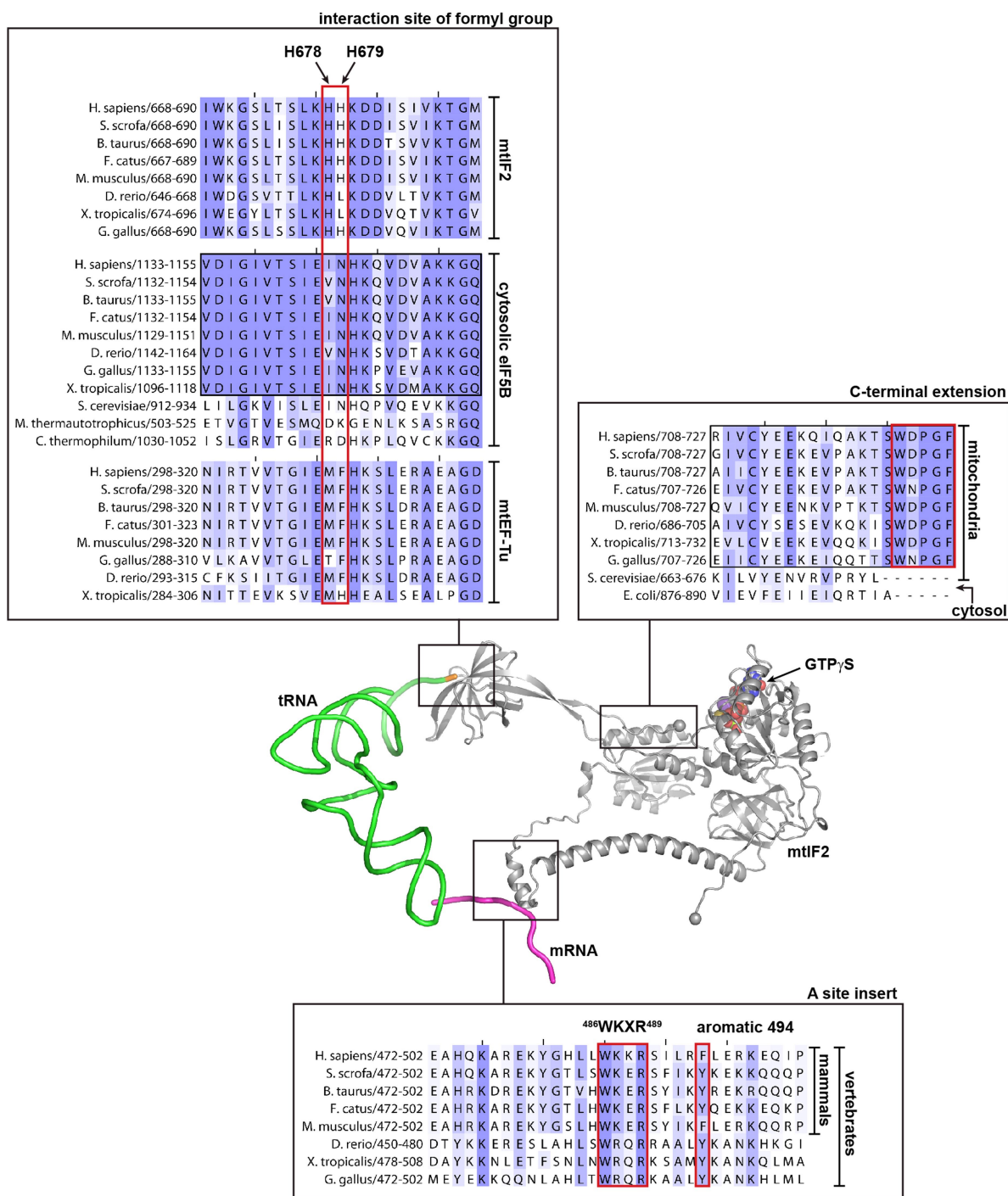
Extended Data Fig. 2 | Map evaluation. **a**, Local resolution estimation performed in RELION yielded maps that were filtered according to the local resolution estimate⁴⁷. Displayed are a front view and a slabbed view at a position indicated by the arrow on the left. Colour keys indicate the local resolution in Å. In contrast to the FSC curves in Extended Data Fig. 2b, local resolution has been estimated and depicted for the entire volume of map 1, map 2 and map 3. Map 2 indicates that the ribosomal subunits exhibit a substantial rotational freedom in the initiation complex leading to a poorly resolved 39S if the alignment is focused on the 28S during refinement. **b**, FSC curves calculated from the two particle half

sets from gold-standard 3D refinement (black) or from model versus map (blue) for all three deposited maps and their corresponding PDBs. **c**, A representative micrograph shows the particle distribution of the 55S initiation complex on cryo-EM grids. **d**, Euler angle distribution of particles included in the final 3D reconstructions are shown using the .build file generated in Relion. Distributions for map 1 and map 3 are displayed. Because map 1 and map 2 were generated from the same particle subset, map 2 distribution is expected to be very similar to the one from map 1 and is therefore not shown.



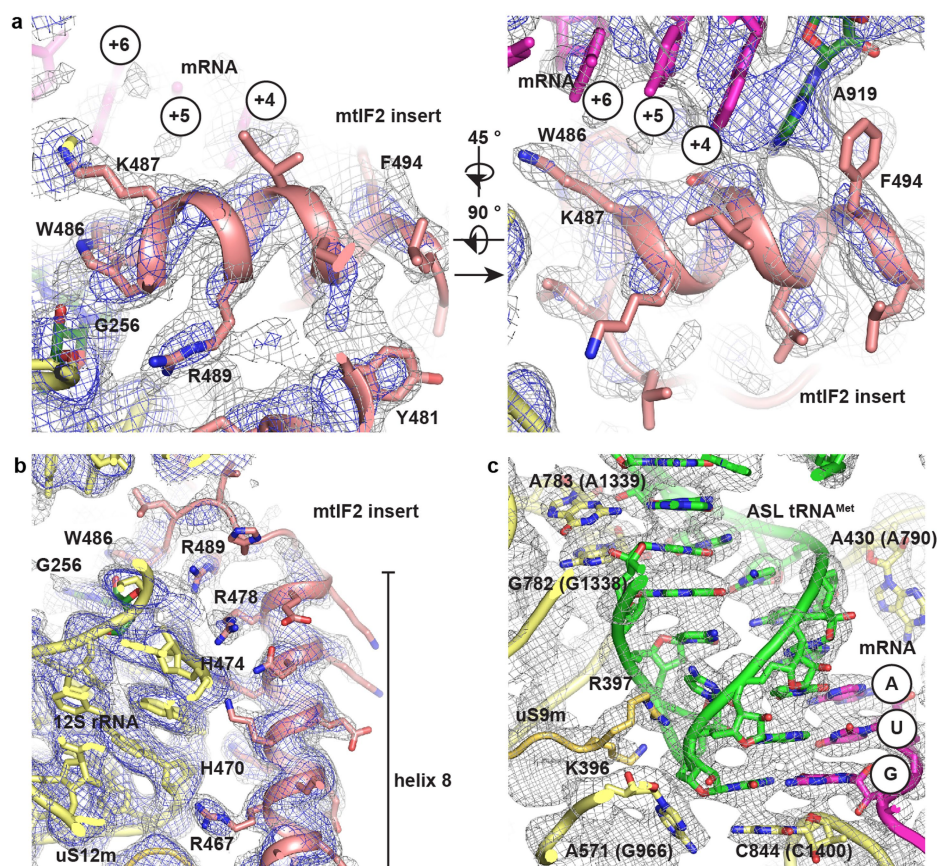
Extended Data Fig. 3 | Quality of the cryo-EM maps around mtIF2 and the role of C-terminal domain of bL12m in subunit joining. **a**, Sliced representation of the ternary complex (red) bound between the ribosomal subunits. All domains of mtIF2 are clearly resolved (maps at 2.2 and 4 σ). **b–e**, Magnified views of different areas of mtIF2. **b**, The contact site of the G domain with domains II and III (maps at 3 and 6 σ). **c**, GTP γ S coordinated by the mtIF2 P loop and switch regions 1 and 2 (maps at 2.75 and 5.5 σ). **d**, Domain IV (maps at 2.75 and 5.5 σ). **e**, Map filtered to 5 Å showing bL12m-CTD bound to the solvent-side of the mtIF2 G-domain and opposite to GTP γ S. The ribosomal L7–L12 stalk forms part of the conserved GTPase activating centre of the ribosomal large subunit in all kingdoms of life⁴⁸. It enhances recruitment of translational GTPases

to cytoplasmic ribosomes via its flexibly attached C-terminal domain (CTD), which has been observed to bind the G' domains in EF-G and RF3^{49–51}. However, G' does not exist in mtIF2. In the initiation complex, we show that bL12m-CTD recognizes the G-domain of mtIF2 on the surface-exposed side opposite the catalytic centre. Since initiation complex formation involves binding of mtIF2 to the small ribosomal subunit before the large subunit joins, the observed interactions of mtIF2 with bL12m may be important to promote subunit joining to form the 55S mitoribosomal initiation complex rather than recruitment of mtIF2 to the assembled ribosome. The bL12m-CTD was modelled with PHYRE2³⁶ using PDB 1CTF³⁸ as a template and docked as a rigid body. Experimental densities are shown at two different contour levels (1.5 and 3 σ).



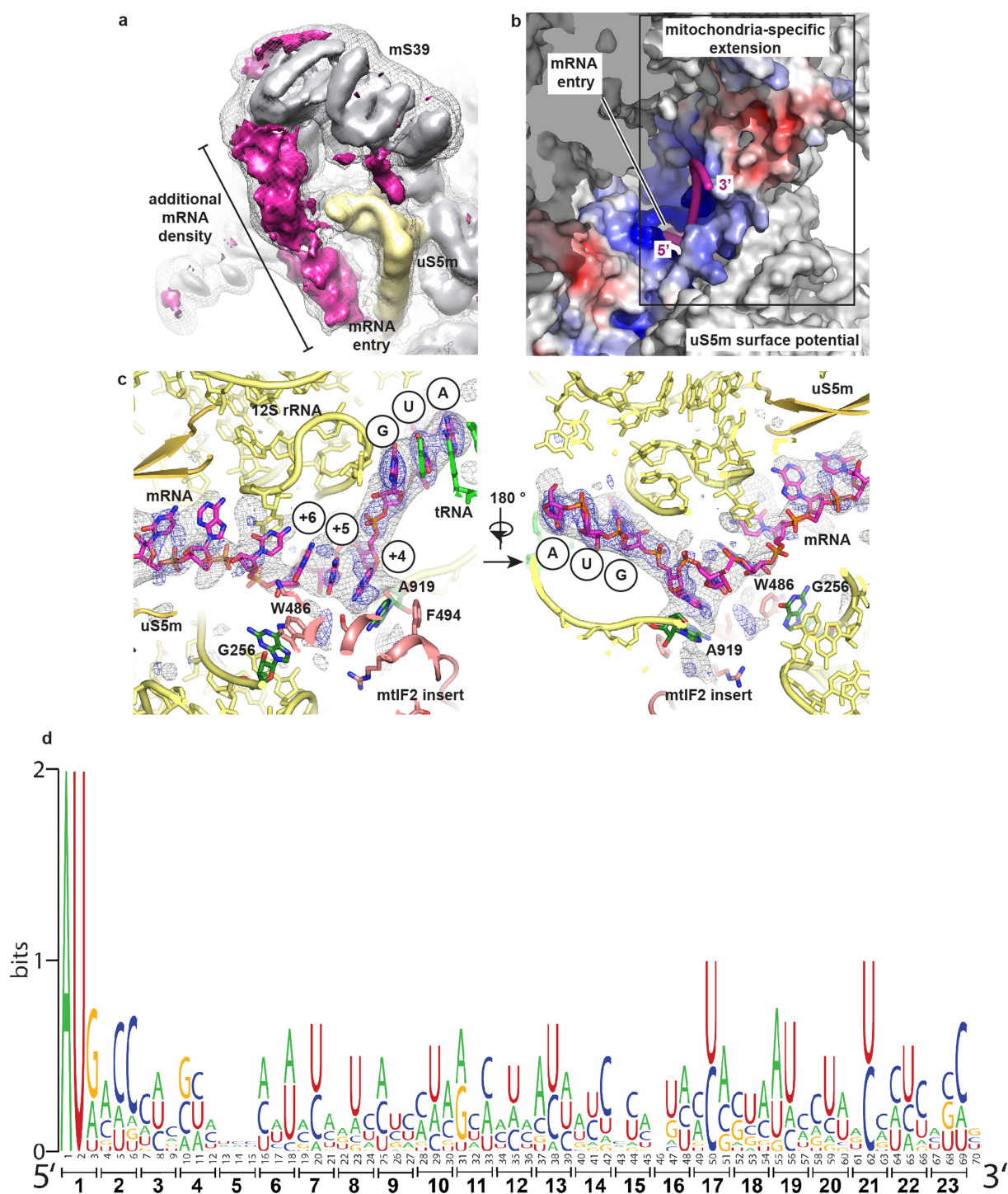
Extended Data Fig. 4 | Sequence alignments for different functionally important regions of mtIF2. Boxes indicate where the aligned sequences are located. The alignments contain sequences from mammals as well as other vertebrates to depict a more general conservation. The alignment for the fMet interaction site also contains eIF5B homologues

for which structures have already been published (*Saccharomyces cerevisiae*¹⁷, *Methanothermobacter thermautotrophicus*³⁹, *Chaetomium thermophilum*⁵²). The alignment for the C-terminal extension of mtIF2 also contains *S. cerevisiae* mtIF2 and cytosolic IF2 from *E. coli*, both of which lack the extension.



Extended Data Fig. 5 | The mtIF2 insert contacts the decoding centre and closes the mRNA channel. **a**, Two views of the α -helical element of the mtIF2 insert occupying the A site with experimental maps at two different contour levels (maps at 3 and 5 σ). W486 stacks on top of decoding nucleotide G256 and F494 contacts the flipped out A919. mRNA bases are numbered according to their position relative to the 5' end of the mRNA. **b**, The mtIF2 insert substantially extends α -helix 8 of the mitochondrial IF2 homologue and then enters the A site. A stable contact is established by a number of conserved positively charged residues facing the 12S rRNA (maps at 3 and 5 σ). **c**, The fMet-tRNA^{Met} anticodon stem loop (ASL) that is in contact with the MT-CO3 AUG start codon is stabilized in the P site by numerous conserved interactions with 12S

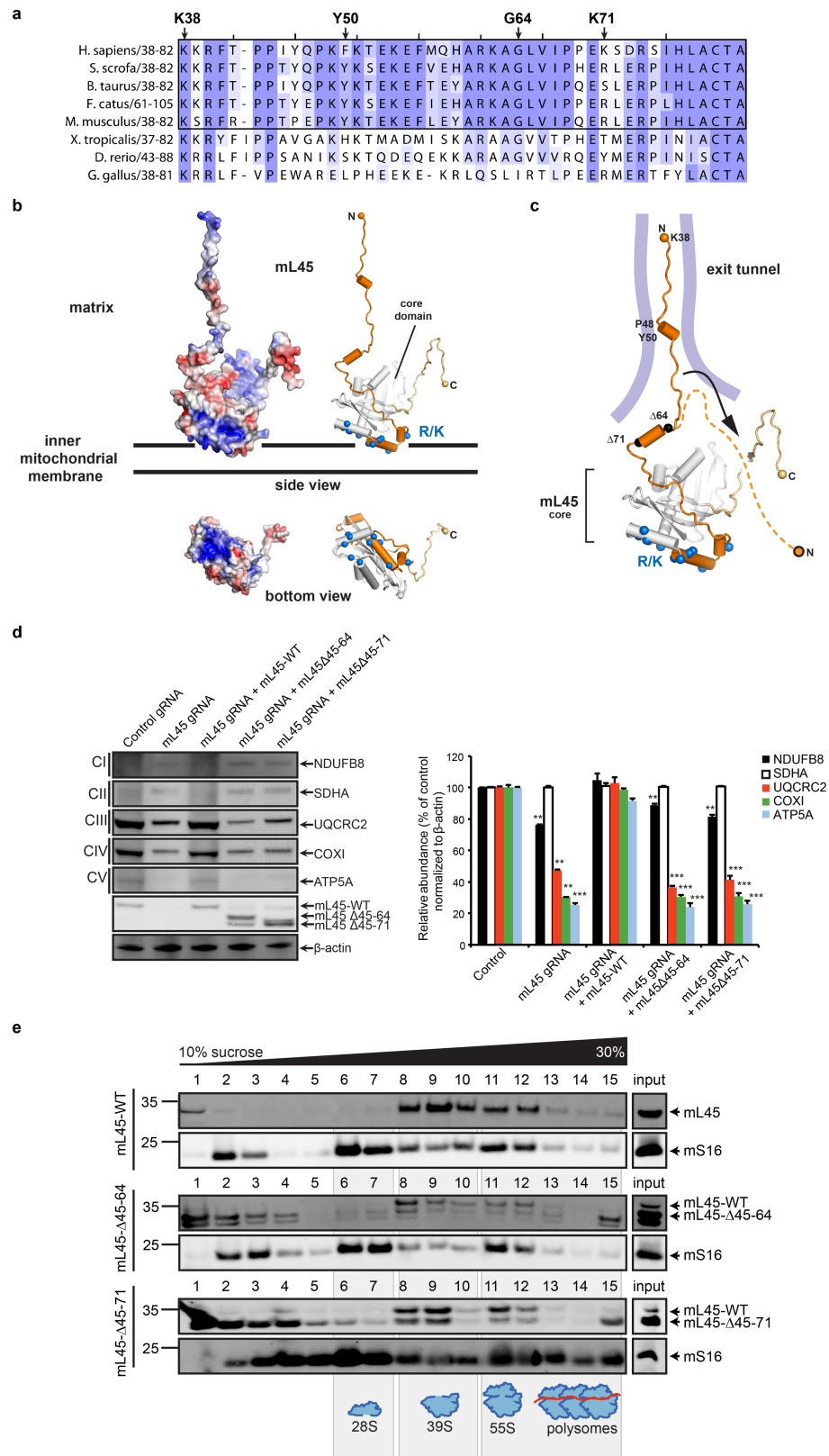
rRNA⁵³. The anticodon fully base-pairs with the start codon, with the tRNA wobble base of the anticodon stacking on top of C844 (C1400 in *T. thermophilus*) and its ribose against A571 (G966). G782/A783 (G1338/A1339) protrude from the 28S head to form A-minor interactions with fMet-tRNA^{Met} specific G-C pairs 26:38 and 27:37. A430 (A790) stacks onto the ribose of tRNA nucleotide 35 to stabilize tRNA binding from the opposite side of the ASL. Protein uS13 is not present in mitoribosomes and thus uS9m is the only ribosomal protein to contact tRNA in the P site. Its C-terminal tail (comprising residues K396 and R397) reaches into a cavity formed by phosphates 30, 31 and 31 of the ASL backbone (map at 5 σ).



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | mRNA binding and start codon selection on the mitochondrial ribosome. **a**, Map 2 (classified for as described in Extended Data Fig. 1 and displayed without post-processing) shown colour coded according to the underlying atomic coordinates (grey, the small ribosomal subunit and mtIF2; yellow, uS5m). A lower contour level of map 2 is shown in transparency and locally filtered in Relion⁴⁷. Density that cannot be assigned to the underlying atomic coordinates reaches from mS39 towards the mRNA entry site (magenta). We believe that this density contains mostly mRNA but it may also include 6 unassigned amino acids from mS39 and possibly part of 21 unassigned amino acids from the N terminus of mS35. **b**, The mRNA entry is surrounded by uS5m. mRNA (magenta) follows the positively charged surface of the mitochondria-specific uS5m extension towards the A site. The surface potential for uS5m was calculated using PDB 2PQR⁵⁴ and visualized with the APBS tool⁵⁵ from PyMOL (M. Lerner and H. Carlson, University of Michigan). ± 5 kT/e electrostatic potential of uS5m have been plotted. **c**, Although resolved to atomic resolution only in the area of the start codon–anticodon

interaction, cryo-EM density for the mRNA can be assigned along its entire path through the mRNA channel, reaching from the P site—where the AUG start codon is located—into the A site, which is shielded by the mtIF2 insert. Subsequently, density nestles alongside protein uS5m, which substantially restricts the diameter of the mRNA channel and places a delineation that may prevent mRNA from slipping out of the mRNA channel. Map 2 is shown unfiltered (blue) and filtered to 5 Å (grey) at two contour levels. For clarity, cryo-EM density for the entire ribosome and mtIF2 has been subtracted from map 2 in Chimera and the difference density is carved 10 Å around our modelled mRNA (contour levels are 10 and 15 σ). The mRNA occupies a similar position as in the elongation complex¹. **d**, Alignment of the first 70 nucleotides of the 11 mRNA 5' ends in human mitochondria (*MT-ND1*, *MT-ND2*, *MT-CO1*, *MT-CO2*, *MT-ATP8*, *MT-CO3*, *MT-ND3*, *MT-ND4L*, *MT-ND5*, *MT-CYB*, *MT-ND6*), starting precisely at the start codon. Codons are indicated by bars and numbered. Alignments were generated using the weblogo server (<https://weblogo.berkeley.edu>).

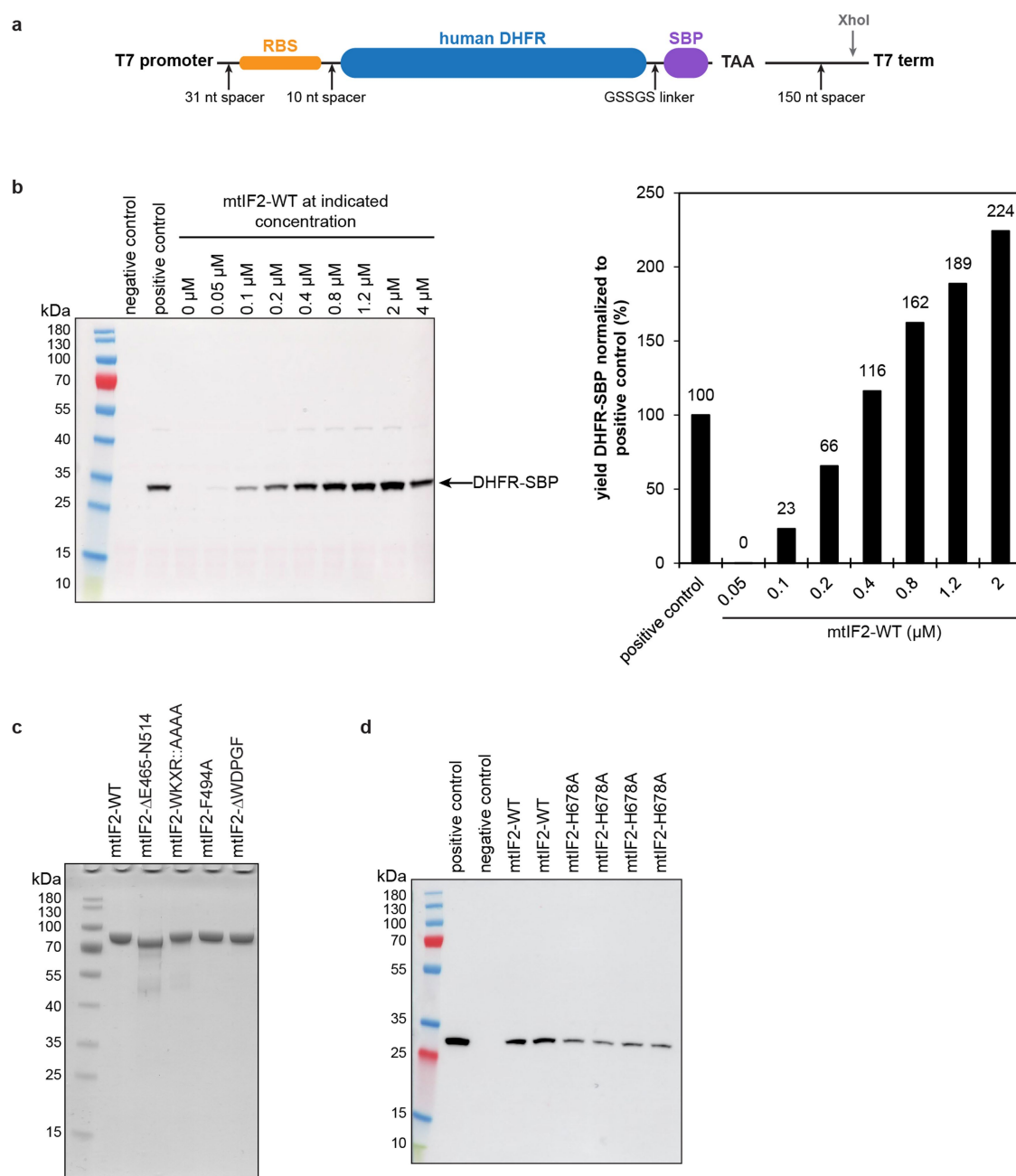


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | mL45 serves as ribosomal membrane anchor and is crucial for insertion of oxidative phosphorylation proteins.

a, Sequence alignment of the mL45 N-terminal extension in vertebrates shows strong sites of conservation—predominantly in mammals. **b**, The surface potential of mL45 (calculated using DELPHI implemented in MOLDRW⁵⁶ and visualized with the APBS tool of PyMOL) shows that the membrane-facing side of mL45 contains a large positively charged patch that may mediate association to the negatively charged inner mitochondrial membrane⁵⁷. For comparison, the structure is shown as a cartoon in the right panel, with positions of positively charged residues in the putative membrane interaction area as blue spheres. **c**, Polypeptide synthesis necessitates displacement of the NTE at a hinge region around G64. The mL45 NTE was truncated at positions G64 and K71 to study its role in vivo. Locations of positively charged residues in mL45 α -helices proposed to mediate membrane association of the ribosome are indicated as blue spheres. **d**, Left, cell lysates (25 μ g) from HEK293T

cells co-transfected with control or mL45 CRISPR/Cas9 plasmids and MRPL45 wild type or deletion mutant expressing plasmids were resolved on 4–20% SDS–PAGE gels and immunoblotted to investigate the steady-state levels of nuclear- and mitochondrial-encoded oxidative phosphorylation (OXPHOS) and ribosomal proteins. β -actin was used as a loading control. The data are representative of at least three independent biological experiments. Right, quantification of the relative abundance of the OXPHOS polypeptides relative to control and normalized to the β -actin loading control. Error bars indicate standard error of the mean. $^{**}P < 0.01$, $^{***}P < 0.001$, Student's *t*-test. **e**, A continuous 10–30% sucrose gradient was used to determine the distribution of the small and large ribosomal subunit and polysomes in mitochondria isolated from cells expressing wild-type or truncated mL45. Mitochondrial ribosomal protein markers of the small (bS16m) and large (mL45) ribosomal subunits were detected by immunoblotting. The input, mitochondrial lysate, was used as a positive control. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 8 | Additional information on in vitro translation assays. **a**, Depiction of the construct used for in vitro translation assays. Human DHFR was fused to streptactin binding protein (SBP) via a linker encoding the amino acid sequence GSSGS. Ribosome binding site (RBS), linker regions (nt, nucleotides) as well as T7 promoter and terminator have been copied from the PURE Express template plasmid provided by the manufacturer (New England Biolabs). The XhoI cleavage site was used to generate DNA templates for run-off transcription. DHFR-SBP was efficiently expressed either upon addition of DNA as template or after supplying DHFR-SBP mRNA directly (data not shown). We decided to perform all subsequent experiments providing equal amounts of mRNA to ensure that every sample contains the same concentration of template in order to make translation yields comparable. **b**, Production of DHFR-SBP was monitored after 2 h at 37 °C at different concentrations of mtIF2(WT). The positive control contained *E. coli* IF1, IF2 and IF3 but no mtIF2. The negative control contained all *E. coli* initiation factors but lacked mRNA.

mtIF2 was tested at the given concentrations and in the presence of *E. coli* IF3. Immunoblots (left) were quantified using the gel analysis routine in ImageJ (right). The sample containing 4 μM mtIF2 was excluded from quantification because it was only partially transferred onto the nitrocellulose membrane during blotting. **c**, SDS-PAGE with 2 μg protein loaded for each mtIF2 variant to show that protein concentrations have been estimated correctly for all variants before in vitro translation was performed (data for mtIF2(H678A) not shown). **d**, Since experiments to determine the translation activity of mtIF2(H678A) were performed at a later time point than for other variants, samples were analysed on separate immunoblots. The immunoblot shows samples from 4 independent experiments for mtIF2(H678A) and 2 independent experiments for mtIF2(WT). Bands have been quantified as for other mtIF2 variants using the gel analysis routine in ImageJ and activity was normalized to mtIF2(WT) (see Fig. 2).

Extended Data Table 1 | Cryo-EM data collection and refinement statistics

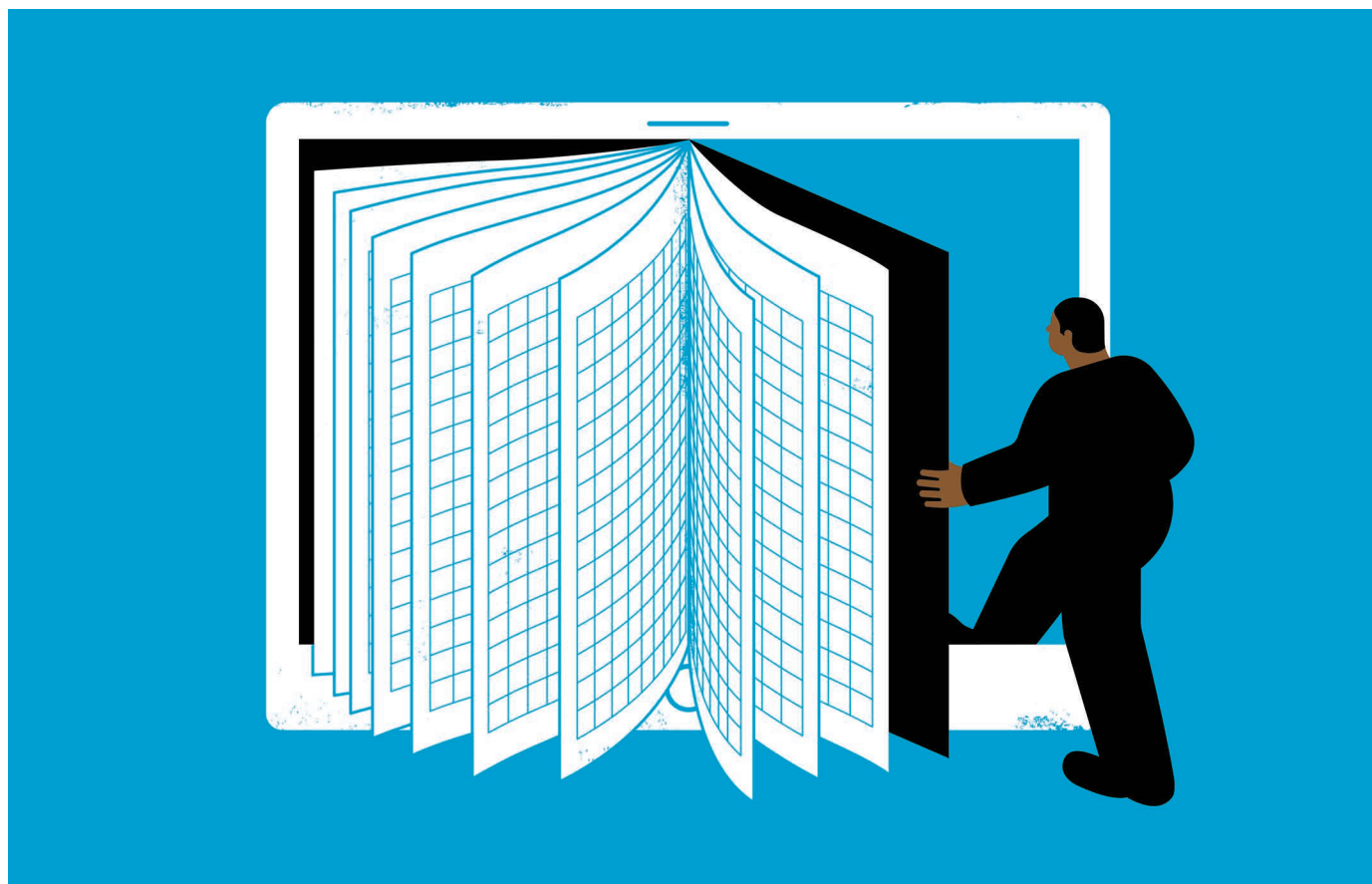
EM Data collection			
Microscope model	FEI Titan Krios		
Detector model	FEI Falcon 3EC		
Number of datasets	3		
Number of micrographs collected	13936		
Magnification	100719x		
Voltage (kV)	300		
Electron dose (e ⁻ Å ⁻²)	40		
Pixel size (Å)	1.39		
Defocus range (μm)	1.2 – 2.5		
Symmetry imposed	none		
Name of 3D-reconstruction	28S subunit ^a	39S subunit ^b	55S initiation complex
EMDB map entry	EMD-4369	EMD-4370	EMD-4368
PDB coordinate entry	PDB 6GAZ	PDB 6GB2	PDB 6GAW
Initial particle images (no.)	1,366,787	1,366,787	1,366,787
Final particles images (no.)	139,206	75,666	139,206
Resolution (Å) (at FSC = 0.143)	3.2	3.1	3.2
Map sharpening B-factor (Å ²)	-151.5	-140.5	-83.2
Reciprocal space refinement and model validation statistics ^c			
Initial model used (PDB code)	4v1a	4v19	6GAZ/6GB2
Model resolution (Å) (at FSC = 0.5)	3.2	3.3	3.2
Map resolution range used for refinement (Å)	40.0-3.13	40.0-3.13	40.0-3.15
Map sharpening B-factor (Å ²)	-151.5	-140.5	-83.2
Spacegroup	P1	P1	P1
a=b=c (Å)	390.59	390.59	390.59
α=β=γ (°)	90	90	90
Number of reflections	4057133	4056927	3972521
Model composition			
Non-hydrogen atoms	73170	111114	178376
Protein residues	6303	9311	15043
RNA residues	1048	1687	2664
Ligands ^d	125	228	347
B factors			
Average	73.8	56.9	66.8
Protein	83.3	64.4	75.4
RNA	52.3	41.6	48.7
Ligands ^d	59.8	30.6	37.6
Working R-factor (%)	24.7	24.4	28.0
wxc weighting factor	1.25	1.25	1.25
R.m.s. deviations			
Bond lengths (Å)	0.008	0.006	0.007
Bond angles (°)	1.098	1.017	0.949
Validation ^c			
MolProbity score	2.56	2.54	2.53
Clashscore	11.4	9.4	10.2
Protein			
Poor rotamers (%)	9.6	10.1	9.2
EMRinger score	2.228102	2.339295	1.940439
Ramachandran plot (%)			
Favored	96.4	95.9	96.0
Allowed	3.5	4.0	3.9
Disallowed	0.1	0.1	0.1
RNA			
Correct sugar puckers (%)	99.6	99.4	99.5

^aDuring refinement of the 28S small subunit, mtIF2, the mRNA and fMet-tRNA^{Met} were included.^bDuring refinement of the 39S large subunit, mtIF2 and fMet-tRNA^{Met} were included.^cThe model was validated using the MolProbity server (<http://molprobity.biochem.duke.edu>)⁴⁶.^dLigands include Mg, Na, Zn, HOH, spermine, GTP-γS, GTP, GMP and formyl-methionine.

LAB NOTEBOOKS GO DIGITAL

A burgeoning array of digital tools is helping researchers to document experiments with ease.

ILLUSTRATION BY THE PROJECT TWINS



BY ROBERTA KWOK

Since at least the 1990s, articles on technology have predicted the imminent, widespread adoption of electronic laboratory notebooks (ELNs) by researchers. It has yet to happen — but more and more scientists are taking the plunge.

One barrier to uptake is the wide range of products available. ELNs comprise software that helps researchers to document experiments, and that often has features such as protocol templates, collaboration tools, support for electronic signatures and the ability to manage the lab inventory. But the ELN market encompasses considerable variety; a study conducted in 2016 by the University of Southampton, UK, identified 72 active products

(S. Kanza *et al.* *J. Cheminformatics* **9**, 31; 2017). “It’s just insane,” says Sian Jones, a petroleum engineer at the Delft University of Technology in the Netherlands. “It does become very confusing.” And many researchers simply lack the time or motivation to make the move to ELNs.

But today’s early-career researchers, who have grown up with digital technology, tend to expect — and to embrace — electronic solutions. Recent trends in research have also created a demand for such changes: as scientists deal with increasing volumes of data, gluing printed results into a paper notebook becomes more archaic. Concerns over reproducibility, as well as more stringent requirements on data management from funding agencies, have motivated improvements in the documentation of lab work. And the ELN

market has expanded to include more intuitive tools, such as cloud-based products, which are easier to adopt than those requiring information technology (IT) support to install. “I do feel that we’re approaching a tipping point,” says Alastair Downie, head of IT at the Gurdon Institute at the University of Cambridge, UK.

ELN developers say that they have also seen signs of growing interest. Where researchers once questioned the utility of ELNs, now they are quicker to commit, says Simon Bungers, co-founder of labfolder, an ELN company in Berlin. Benchling, an electronic research platform in San Francisco, California, has seen use of its ELN in academia more than double for the past two years, with tens of thousands of researchers now logging in every day, says chief executive Sajith Wickramasekara. ►

► And many universities have started to provide such products to their researchers. For instance, LabArchives in Carlsbad, California, has sold campus-wide site licences for its ELN platform to more than 375 research institutions worldwide. (Last month, LabArchives announced a partnership with Macmillan Learning of New York City, which is part of Holtzbrinck Publishing Group in Stuttgart, Germany; Holtzbrinck is the majority shareholder in *Nature's* publisher, Springer Nature.)

Advocates tout the many advantages of ELNs over their paper counterparts. They are easy to search, copy and archive. And thanks to templates, scientists don't have to rewrite protocols. Researchers can link experiments to specific samples or files, as well as share information easily with other lab members and collaborators, facilitating reproducibility. And supervisors can monitor the activity of their teams remotely.

But there are downsides, too. Although many companies offer free versions of their ELN software, those often come with limits on the number of users, data storage or file size. If the company folds or raises its prices, researchers might find themselves with only a PDF export of their data, which they are then unable to transfer to a competing product. Network interruptions could temporarily restrict access to data. And researchers might still prefer to make some notes or sketches on paper at the bench, which must then be imported into the ELN.

Despite these shortcomings, more and more researchers are going digital. To find a software solution that suits your needs, experienced users suggest taking the following steps.

Get educated. Online resources can give prospective users a sense of the market. Downie's guide to ELNs (go.nature.com/2v7iayq), hosted on the Gurdon Institute's website, includes information on attributes such as cost tiers, support for computing platforms, and where the data can be stored for 28 products. The Electronic Lab Notebook Matrix (go.nature.com/2n54fma), collated by Harvard Medical School in Boston, Massachusetts, lists the details of more than 50 features for 27 ELNs. And labfolder provides a guide to 16 popular ELNs (go.nature.com/2vco2hz).

Calculate costs. Paid versions of most ELN services used in academia cost US\$10–20 per user per month, Downie says. The restrictions that are associated with free versions of these tools might be malleable, particularly as storage prices fall; Wickramasekara says that the 10-gigabyte limit on Benchling's free academic platform, for instance, can often be raised on request. Open-source options such as the Open Science Framework from the Center for Open Science in Charlottesville, Virginia, also are available.

Understand legal issues. Some funders place restrictions on where data can be stored, so researchers should keep this in mind when

evaluating cloud-based ELNs. Scientists who use personal data that fall within the scope of the European Union's General Data Protection Regulation should consider whether an ELN's data storage complies with those rules. Choosing ELN software that enables completed

“You can't just stick your toe in the water. You've got to dive all the way in.”

pages to be locked and electronically signed could be crucial if the documents are needed to defend researchers against claims of fraud, or must be submitted to the US Food and Drug Administration as part of regulatory processes. Digitally signed and witnessed documents could also be used as evidence in a patent dispute, says Denise Callihan, who manages library services, including patent searching and ELN system, for paints and coatings company PPG in Monroeville, Pennsylvania. PPG uses an ELN software called PatentSafe from Amphora Research Systems in Andover, Massachusetts.

Evaluate stability. Researchers might want to assess the ELN company's chances of survival. Daureen Nesdill, a research-data-management librarian at the University of Utah in Salt Lake City, says she considered this question when evaluating options in 2010. She favoured LabArchives, partly because the company's executives had already launched successful bibliographic-management software. Nesdill advises researchers to choose a company that is at least five years old, has stable funding and states in its terms of service that users will be able to access their data if the firm goes under or is sold.

Think mobile. Some labs prefer ELNs that can run on mobile devices. That was the case for Richard Gates, a chemical engineer at the US National Institute of Standards and Technology in Gaithersburg, Maryland. He and his colleagues wanted to use tablets to record experiments while working in a clean room, because the devices are portable and can be wiped down easily. The researchers, who chose Microsoft's note-taking software OneNote as an ELN, use the tablet's camera to take photographs of instruments and results, and a stylus to annotate images.

Consider software integration. Links to favourite software could tip the scales for some scientists. Organic chemists, for instance, might prefer the PerkinElmer Signals Notebook from PerkinElmer in Waltham, Massachusetts, says Nesdill, because it integrates with the company's chemical-structure-drawing software ChemDraw, enabling structures to be added to the ELN. ResearchSpace in Edinburgh, UK, integrates its ELN with tools such as software-development platform GitHub and reference manager Mendeley, Jones notes.

Go for a test drive. Jones suggests test-driving free versions of a few products, ranging from basic to complex. “Don't look at more than four, otherwise your head explodes,” she says. While evaluating several ELNs last year, Christoph Seiler, who runs a facility for zebrafish experiments at Children's Hospital of Philadelphia in Pennsylvania, asked himself, “Is that an interface I can use every day?” He settled on Benchling, partly because he found its ELN to be attractive and well-organized.

Preferences for minor features come down to personal taste. For instance, Downie likes the way that the ELN from SciNote in Middleton, Wisconsin, provides a flexible, flow-chart-like structure, and Jones enjoyed seeing a feed of other users' activities in Labguru, an ELN from BioData in Cambridge, Massachusetts. (Digital Science in London, which is part of Holtzbrinck, is an investor in BioData.)

Try generic platforms. Some scientists stick with generic note-taking products. Michael Gotthardt, a cardiovascular-disease researcher at the Max Delbrück Center for Molecular Medicine in the Helmholtz Association in Berlin, chose OneNote because he wanted a low-cost product with “essentially no learning curve” that the IT department could install locally with ease. Every month, his team exports pages to PDF files and signs them electronically; the files are then moved to a directory where they cannot be changed. Evernote, from Evernote Corporation in Redwood City, California, is an alternative note-taking option.

Commit to change. In 2017, Downie co-led a trial of four ELNs, in which researchers at the University of Cambridge rated features such as user interface, support for collaboration and file-management capabilities. Although many scientists initially expressed enthusiasm about ELNs, only 37 of the 161 participants completed the exercise. “It shows the level of commitment that's required,” Downie says. “You can't just stick your toe in the water. You've got to dive all the way in.”

That said, some acclimatization might be required. Gotthardt gave his team three months to play with OneNote while continuing to record experiments on paper. Everyone then made the switch — a change that went smoothly, he says. Ulrich Dirnagl, an experimental neurologist at the Berlin Institute of Health, which provides labfolder to employees at one of its institutions, says that he has seen the most uptake when one lab member starts using an ELN and word spreads to colleagues, rather than when the entire group is forced to convert.

“Before, they said, ‘I don't need this, and I just want to scribble down my little notes,’” Dirnagl says. “Three weeks into the ELN, they want to press a button for a cappuccino.” ■

Robert Kwok is a freelance science writer based in Kirkland, Washington.

CAREERS

EARTH HACKER: Harnessing big data to leverage impact **p.273**

BLOG Personal stories and careers counsel
blogs.nature.com/naturejobs

NATUREJOBS For the latest career listings and advice www.naturejobs.com

MICHAEL GOTTSCHALK/PHOTOTHEK/GETTY



Get your science into the news with help from your press office.

Press ahead

Public information officers can help scientists to share their research more widely.

BY ROBERTA KWOK

Megan Thoemmes knows first-hand that a good press officer can catapult scientific discoveries into the media spotlight.

She and her colleagues were about to publish a paper in mid-2014 about mites that live on human faces¹. Sensing a story that could catch the public's interest, her adviser notified Matt Shipman, the research-communications lead at the press office at North Carolina State University in Raleigh, where Thoemmes is a PhD student in ecology and evolutionary biology. Shipman edited a blogpost on the study for the university's website and coached her for media interviews.

That media training came in handy. After Shipman notified journalists, outlets such as *National Geographic*, US-based NPR (National Public Radio), *Wired* and Radio New Zealand

contacted Thoemmes for comment. The coverage was positive and mostly accurate, she says. The process taught her how to distil results into a few key points, among other skills. And other benefits emerged: a medical researcher noticed the news and initiated a collaboration with her team on how mites influence the microbiome.

This year, Shipman wrote a press release about another paper by Thoemmes, published in May, on the communities of microbes and arthropods living in chimpanzee beds². For the next two weeks, Thoemmes was deluged with interview requests from *The Washington Post*, the BBC, UK online newspaper *The Independent* and W Radio in Colombia, among others. She credits Shipman with helping her to reach a global audience. Without a press release, "I absolutely would not have gotten that amount of coverage," she says.

Many researchers do not take advantage of their institution's press office, perhaps because they feel they lack the time for media outreach or are dubious about the benefits. But those who do reach out often find that press officers help them to craft a clear message, connect them with journalists (see 'On the record'), and increase the visibility of their research. Press officers say that they have seen such publicity bring career benefits such as collaborators, graduate students, work opportunities and attention from funders.

However, researchers need to ensure that press releases do not hype findings, say science-communication experts. In today's struggling journalism industry, many news outlets lack the resources for thorough, sceptical reporting of science news, and content-aggregator websites reprint press releases almost word-for-word. "It certainly increases the responsibility of the press officers and the scientists who issue these press releases to be balanced and responsible and cautious," says Marina Joubert, a science-communication researcher at Stellenbosch University in South Africa, and a former freelance science communicator. In addition, not all press releases will prompt major news coverage — and some might provoke criticism of the research.

Many scientists promote their work on their own through social-networking sites, such as Twitter and Facebook. But few can match the huge audiences that the leading mass-media outlets command. For instance, *The New York Times* has more than 2.6 million digital-only subscribers, and BBC News reaches about 347 million people per week. "There's nothing else that has the reach of media," says Jonathan Wood, communications manager at the Francis Crick Institute in London.

News articles about research can bring direct career benefits. For instance, Wood has heard from scientists at other institutions where he has worked that media coverage encouraged funders' interest in their research. When Joubert provided freelance science-communication services to the University of Pretoria in Hatfield, South Africa, she wrote a press release about a soil study. It led an organization to invite the researcher to apply for a work contract at a national park; he won the contract and accepted. And Shipman points to studies that have found correlations between news coverage and higher citation rates^{3,4}, and between media interactions and higher h-indices⁵ — the latter being a measure of the impact of an author's body of research. "There are a number of very selfish, practical ►

► reasons for researchers to engage with their press office,” he says.

For early-career scientists seeking academic jobs, media coverage could provide a boost. Dawn Levy, lead science writer for Oak Ridge National Laboratory’s physical-sciences directorate in Tennessee, notes that she has seen publicity help researchers to garner award nominations, speaking invitations, funding offers and research partnerships. Although she is sceptical that news articles lead directly to more citations, she says that media coverage — and its explanation of why a research project matters — conveys to the reader, viewer or listener that the scientists are trying to solve a significant problem. “That’s a great reason for someone to hire you,” Levy says, “regardless of citations.”

News coverage of geoscientist Simon Cook’s study of glaciers⁶ in Bolivia helped him to stand out when he was applying for academic jobs in 2016, he suspects. The paper, published in *The Cryosphere*, was press-released by the European Geosciences Union (EGU), which publishes the journal, resulting in coverage by major media outlets. Cook, now at the University of Dundee, UK, advises applicants not to focus solely on news coverage. Ultimately, he says, landing a position requires consistently high-quality research, whether it earns press attention or not.

And not every press office has the skills or resources to make a media splash. An inexperienced press officer could have fewer connections to journalists or might use ineffective techniques such as generic e-mails to reporters instead of targeted pitches, Joubert says. At smaller institutions, the press office might consist of a single person who has limited time to spend on each story. If researchers find that their press office cannot offer enough help, they could consider hiring a freelance science communicator, Joubert says. Scientists can find such freelancers by attending science-communication events or asking regional or national science-communication associations.

DISTILL THE MESSAGE

Scientists should consider notifying their press office of a forthcoming paper if they think it will have a strong impact in their field, has practical applications or simply is cool, Shipman says. Other possible newsworthy topics include a conference presentation, award, grant or clinical trial. Typically, the junior or senior researchers on a team can alert the press office by e-mail or phone. Timing recommendations vary; some press officers prefer that researchers let them know about a paper as early as submission, while others say that within a week or two of acceptance usually still allows enough time to put out a press release.

To promote a paper, an institutional press office can collaborate with the journal’s press office to reach more reporters, says Bárbara Ferreira, media and communications manager at the EGU, who publicizes papers

ON THE RECORD

Media management

Media outreach can be intimidating for scientists. Here are some tips to navigate interviews with journalists.

- After the press release is issued, be available for interviews for about one week.
- Return journalists’ calls and e-mails promptly. Reporters often face extremely tight deadlines.
- Prepare a few key points.
- Be conversational — don’t be afraid to show a human side and tell anecdotes.
- Assume that everything you say can be published.
- Most journalists will not send you a draft of the article to look over. However, researchers can offer to answer fact-checking questions.
- If an article contains an error or sensationalizes the findings, ask for a correction, write a letter to the editor, or respond on social media and tag the journalist. **R.K.**

published in the union’s 17 peer-reviewed journals. For example, a university press officer in Germany might have mostly German media contacts, but Ferreira can send the press release to a larger network of European journalists.

If it’s too late to issue a press release, scientists can still spread the message about their work. For example, they could ask a press officer to write a feature about a study for their institution’s website or promote the researcher’s blog post on social media, Levy says. Some press officers set up ‘Ask Me Anything’ sessions on Reddit, a news-aggregator site and forum, which allow researchers to answer questions from the public.

If the work is timely enough for a press release, the press officer typically interviews

the scientist about points such as the motivation for the study, how it builds on previous research, the most surprising or interesting results, and possible societal implications. When reviewing the draft of the press release, researchers should correct inaccurate or misleading sentences and express concerns about any statements that make them uncomfortable, but they should avoid adding technical details and jargon.

In particular, the title should be short and snappy because it is often used as the subject line in e-mails to reporters, Ferreira says.

Scientists should ensure that findings and implications are not hyped. In a 2014 study, researchers analysed health-related press releases from UK universities and found that 33–40% contained statements that went beyond the scientific paper, such as making a stronger causal link between two factors than had been stated in a correlational study⁷. The results suggest that journalists are not the only ones to blame for exaggerations, says study co-author Petroc Sumner, a psychologist at Cardiff University, UK. “A good percentage of them are already there in press releases,” he says.

Sumner also advises scientists to use hedge words such as ‘may’, ‘might’, or ‘could’ when describing correlational evidence and to add caveats. Importantly, his team found no evidence that including caveats in press releases reduced news coverage.

Press officers can help to reduce the risk of misinterpretation by explicitly stating what the research does not imply. Climate scientist Carl-Friedrich Schleussner had this experience when working with the press office at the Potsdam Institute for Climate Impact Research in Germany.

His team’s study suggested that climate-related disasters increased the risk of armed conflict in countries with multi-ethnic societies⁸. To avoid implying a causal link, the press release included a quote from Schleussner stating, “Climate disasters are not directly triggering conflict outbreak.” While this statement did not prevent reporters from asking him



SIMON COOK

Media coverage of his glacier research might have helped Simon Cook to stand out when applying for jobs.

whether climate change had caused the Syrian civil war or the influx of refugees to Europe, working out the phrasing in the press release beforehand helped him to stick to a firm message during interviews, says Schleussner, who works at the non-profit organization Climate Analytics in Berlin. The media coverage was generally accurate, he says.

Scientists should consider who might react negatively to the news, Joubert says. For instance, the press release that she wrote about soil research suggested that four-wheel-drive vehicles (also called 4x4s) damaged the environment and should not be allowed off-road in protected areas. Members of 4x4 clubs complained online and by e-mail. In retrospect, she says, she could have made it clearer in the headline that the suggested ban applied only to off-road driving and that many 4x4 drivers have important roles in nature conservation, such as supporting national parks.

Images and videos are important elements of press releases. Scientists should supply pictures that are not in the paper because the journal typically owns the copyright to the paper's images, Levy says.

She advises using pictures that do not contain text, and adds that videos should be under one minute long; the press office can take care of posting videos on YouTube. Press offices at large institutions may include a photographer, videographer or graphic designer who can produce multimedia output as well.

The press officer could ask the scientist to suggest publications that should be notified about the research. Although press officers are well-versed in mainstream media outlets, researchers might be aware of niche publications read by colleagues. Press releases are often issued through websites such as EurekAlert!, Newswise and AlphaGalileo, and the press officer might also email targeted pitches to journalists.

With a press officer's help, researchers can reach more people outside their field. "This really is your chance to explain why your work is interesting and important," Wood says. "After all, you're the expert. Who better for us, the public, to hear from?" ■

Roberta Kwok is a freelance science journalist in Kirkland, Washington.

1. Thoemmes, M. S., Fergus, D. J., Urban, J., Trautwein, M. & Dunn, R. R. *PLoS ONE* **9**, e106265 (2014).
2. Thoemmes, M. S. *et al. R. Soc. Open Sci.* **5**, 180382 (2018).
3. Kiernan, V. *Sci. Comm.* **25**, 3–13 (2003).
4. Phillips, D. P., Kanter, E. J., Bednarczyk, B. & Tastad, P. L. *N. Engl. J. Med.* **325**, 1180–1183 (1991).
5. Liang, X. *et al. Journalism Mass Comm.* **91**, 772–791 (2014).
6. Cook, S. J., Kougkoulos, I., Edwards, L. A., Dortch, J. & Hoffmann, D. *Cryosphere* **10**, 2399–2413 (2016).
7. Sumner, P. *et al. BMJ* **349**, g7015 (2014).
8. Schleussner, C.-F., Donges, J. F., Donner, R. V. & Schellnhuber, H. J. *Proc. Natl Acad. Sci. USA* **113**, 9216–9221 (2016).

TURNING POINT

Earth hacker

As a PhD student in 2006, Lucas Joppa launched his academic career in ecological theory. But after starting a postdoc at Microsoft in 2010, becoming chief environmental scientist in 2017 and chief environmental officer this July, he now develops tools that harness the power of big data to inform decision-making in the field.

What are your professional goals?

I want to find answers to big existential questions — such as how humans impact ecosystems and how that ultimately affects life on Earth. These questions have become increasingly computational as we amass both huge data sets from all over the world and the tools to analyse that volume of data. I wanted to work with the people inventing the best techniques to sift through all these data — methods that can help to solve the world's most pressing challenges, from maximizing crop production to tracking endangered animals.

Describe Microsoft at the time of your arrival.

I moved to the company in Cambridge, UK, in 2010. People at the forefront of their fields — from machine learning to theoretical mathematics — all work on common problems together. Everybody has the intellectual curiosity to care a lot about everybody else's work.

What did your advisers think of your postdoc?

Some thought it was risky. A member of my PhD committee, a leading ecologist at Duke University in Durham, North Carolina, took me to lunch and asked if I was certain that I wanted to do this. I'm sure it looked as if I had jumped overboard without a life jacket. But I thought it was the safest path.

Were you a computer geek growing up?

No. I grew up in north Wisconsin with no television. With my undergraduate degree in wildlife ecology, I thought that I'd be a game warden.

When did you realize that ecology needed more computational power?

Day one of my PhD. Everywhere I looked, there was no way around it. I was working on ecological networks — from predator–prey relationships to plants and pollinators. You can't resort to pen and paper when you are researching extinctions in complex networks, or when you are using global satellite imagery to determine



whether deforestation has encroached on 150,000 protected areas.

How did your role evolve into that of chief environmental scientist?

A couple of years ago, my mentors at Microsoft urged me to write down how the organization could leverage its investments in computational research to address issues related to the environment, conservation and sustainability. I wrote a memo, called AI for Earth, which detailed the prospects for using artificial intelligence (AI) to improve environmental sustainability. It was published as a Comment (L. N. Joppa *Nature* **552**, 325–328; 2017) at around the time I took on a more-corporate role. To our knowledge, this position is a first for the technology industry.

Describe Microsoft's AI for Earth programme.

When we announced the US\$50-million, 5-year investment for tackling global environmental challenges in December 2017, most people were focused on the dollar amount. But I am struck by the five-year commitment — that's a geological age for the tech sector. It gives me the stability to form partnerships, award grants and foster research to find ways to protect biodiversity and identify crucial areas for conservation.

Do you have advice for ecologists who want to use AI and computing power?

Don't wait. Anyone can get started with programming languages such as R or Python. AI for Earth is about as easy as it can be for PhD students. And we offer small seed grants that require only a one-page form to allow scientists to access Microsoft's best AI technology. Currently, we have 112 grant recipients in 27 countries. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for clarity and length.

HOME CYGNUS

Time to say goodbye.

BY S. R. ALGERNON

Scheduled departure in four hours, said the computer in the wall, breaking the silence.

Dorothy set the translucent plastic box down on her bed and watched it, as if watching would make the inside bigger. Surely, Mission Control could have managed more. A thousand cubic centimetres was too little space for a life. Her quarters on the Titan habitat at least had enough room for her clarinet and her hard-cover edition of Audubon's *Birds of America*.

The box at that moment contained two pairs of earrings, a necklace and a plastic trinket that generated a hologram of the Valles Marineris canyons when you pushed a button on the top. The hologram projector took up too much space, but it was a family heirloom like the rest. Her father had given it to her at the spaceport before heading to Mars.

Dorothy glanced again at Titan through the window. *If 16 Cygni Bb is as hostile as the recruitment blurb says, I'll be spending most of my time watching the world through screens and windows, watching it freeze and burn.*

Dorothy felt like a prop master on a movie set. Somewhere in this prefab habitat room was the perfect prop, something that would open that window and bring the planet to life for her — and maybe for the Cygnans who followed her — but what was it? Was it a pair of slippers to whisk them back home? Was it a shattered snow globe to unleash a storm of loss and regret? Was it a falcon, shiny and treasured but ultimately meaningless?

She wished she had room for the weather vane on Great-Grandpa's old farmhouse. It was cast iron; it would survive the Cygnan summer, even if it glowed a little or disappeared within a blizzard from time to time.

Mission Control said that we had to give up on Earth's gods, and that they could not protect us, but something is guiding me, even if I can't put a name on it.

"Computer," said Dorothy. "Place a call to Great-Grandpa. Try the landline." The farmhouse was

a billion miles away, but she knew Great-Grandpa would be there. Where else would he be?

"It's me. Dot," she said, over the faint undercurrent of line noise. "I know this message will take an hour to reach you, and then an hour to get back, so I'm just going to talk. I'm not coming back this time, and

chimes and the porch swing, to the rain on the roof and the melody of a Western meadowlark. Every so often, Great-Grandpa would interrupt the natural rhythm with "Did I ever tell you about the time when ..."

Scheduled departure in one hour.

Dorothy wished she had more to say. Did Great-Grandpa ever want to listen to her stories?

It had always been the other way around. Still, she talked about whatever came to mind, about spaceports, exoplanets and all the plans that had fallen through. Her packing took on an urgency now. She downloaded movies — home movies and Hollywood classics — to a removable drive. The rest of the clutter no longer mattered. Mission Control would sell it off or station security would haul it away.

Scheduled departure imminent. Proceed to security checkpoint.

Already?

"I know, Computer. Hush."

Wait. Not just yet. One last thing.

"Great-Grandpa," said Dorothy. "I know you're still listening, so I just want to ask you, are you happy for me? Are you happy for all of us? Should we have stayed? Did we go in the direction you wanted?"

Dorothy knew she would not live long enough to really answer that question for herself, but she wanted to hear his answer, in his voice.

"I'm sending the coordinates for 16 Cygnus B, Great-Grandpa. If you can find a way to transmit in that direction, keep talking. I'll listen through the ship's comm system as long as I can, until the Doppler shift takes the signal out of range."

At the last moment before leaving her quarters, Dorothy downloaded the recording, took out the drive and put it in the box.

I know what this box is, thought Dorothy. It's a piano in a gin joint, somewhere in the desert, off the stage for a while but not forgotten, and waiting for the next set of hands to bring its old sounds back to life. With resignation, nostalgia and traces of hope, she closed the lid, took one last look around, and stepped through the door. ■

S. R. Algernon studied fiction writing and biology, among other things, at the University of North Carolina at Chapel Hill. He currently lives in Singapore.

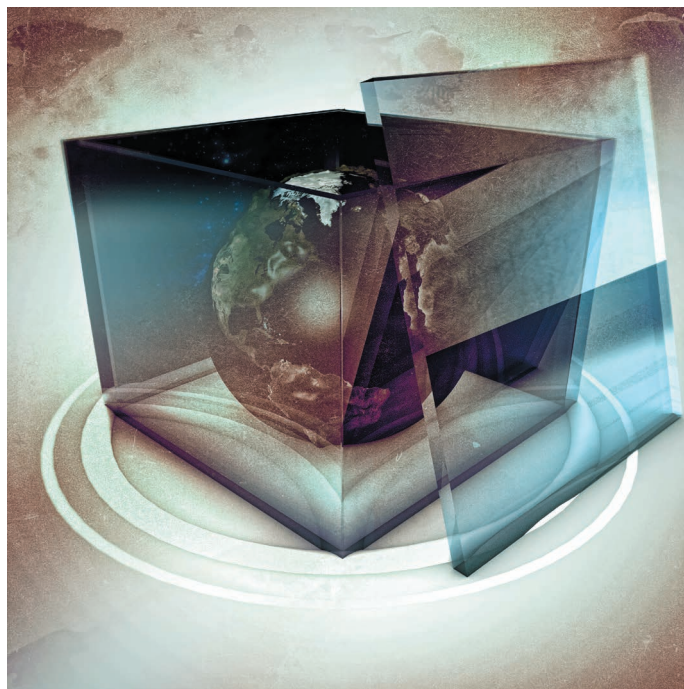


ILLUSTRATION BY JACEY

I might be going far enough away that we won't be able to talk anymore. I know this is sudden, but sometimes life doesn't give you easy choices. I just wanted to thank you, and to hear you one more time."

Dorothy tried to pack, but nothing seemed quite good enough. After two and a half hours, the box was no closer to being full.

Two hours and thirty six minutes later, the sound of wind chimes and creaking wood broke the silence.

"I always knew you'd be the one who wouldn't come back. You're too much like your father. I've never been one for talking much, and not quite so many people to talk to around here. I'll leave the line open as long as you need."

A gust of wind gathered momentum in the background.

"Looks like we're in for a storm," added Great-Grandpa.

You have no idea.

For two hours, Dorothy listened to the